University of Bath

PHD

The evaluation of consolidated bioprocessing as a strategy for production of fuels and chemicals from lignocellulose

Hussein, Ali

Award date:
2016

Awarding institution:
University of Bath

Link to publication

Alternative formats
If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (https://creativecommons.org/licenses/by-nc-nd/4.0/). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy
If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Download date: 04. May. 2023

The evaluation of consolidated bioprocessing as a strategy for production of fuels and chemicals from lignocellulose

Ali H. Hussein

A thesis submitted for the degree of Doctor of Philosophy

University of Bath

Department of Biology and Biochemistry

September 2015

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with the author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

ABSTRACT

Cellulosic biomass is one of the most abundant industrial waste products and an appealing substrate for biorefining strategies to produce biofuels by fermentation. The metabolic engineering of fermentative bacteria, such as the thermophile *Geobacillus thermoglucosidasius*, for high bioethanol yield is well characterised. This has been traditionally facilitated by an economically inefficient multistep process referred to as separate hydrolysis and fermentation (SHF), in which the enzymatic hydrolysis of the cellulosic substrate and fermentation of the liberated sugars is performed sequentially.

Consolidated bioprocessing (CBP) involves performing these two process steps simultaneously, by either introducing cellulolytic capabilities into naturally fermentative organisms or implementing fermentative capabilities in cellulolytic organisms through metabolic engineering. CBP is believed to be a potentially cost-efficient and commercially viable way to produce cellulosic biofuels since the feedback inhibition of glycosyl hydrolases by monosaccharides as they are released is reduced by their rapid conversion through microbial fermentation. This results in faster rates of production and higher yields than those possible with SHF. Furthermore, CBP offers energy savings by removing the need for a complex multistage process with multiple heating and cooling steps.

The aim of the present project is the engineering of CBP capabilities in the ethanologen *G. thermoglucosidasius* through the heterologous secretion of active glycosyl hydrolases into the extracellular milieu or employing surface-layer homology domains to attach them to the bacterial cell-wall. Iterative optimisation will serve to evaluate the feasibility of CBP as a strategy for production of biofuels from lignocellulose using *G. thermoglucosidasius*.

List of Figures and Tables

Table of Abbreviations

ABC	ATP-binding cassette
AddAB	ATP-dependent deoxyribonuclease A/B
AldDH	Acetaldehyde dehydrogenase
AU	Absorbance Unit
BLAST	Basic Local Alignment Search Tool
BSA	Bovine Serum Albumin
C. clariflavum	*Clostridium clariflavum*
C. saccharolyticus	*Caldicellulosiruptor saccharolyticus*
C. thermocellum	*Clostridium thermocellum*
C5	Pentose sugars
C6	Hexose sugars
CAZy	Carbohydrate-Active enZYmes Database
CBM	Carbohydrate Binding Module
CBM	Carbohydrate Binding Modules
CBP	Consolidated Bioprocessing
CBP	Consolidated bioprocessing
***Cc*YoaJ**	*C. clariflavium* DSM19732 expansin-like protein YoaJ
***Cs*cel48/9A**	*C. saccharolyticus* DSM8903 dual exo-endoglucanase A
***Cs*cel5**	*C. saccharolyticus* DSM8903 endoglucanase 5
***Cs*cel5SLH**	*C. saccharolyticus* DSM8903 S-layer endoglucanase 5
***Ct*cel48S**	*C. thermocellum* DSM1237 exoglucanase S
***Ct*cel48Y**	*C. thermocellum* DSM1237 exoglucanase Y
***Ct*cel8A**	*C. thermocellum* DSM1237 endoglucanase A
***Ct*cel9D**	*C. thermocellum* DSM1237 endoglucanase D
***Ct*cel9I**	*C. thermocellum* DSM1237 processive endoglucanase I
***Ct*PinA**	*C. thermocellum* DSM1237 endoglucanase A
DNS	3 5-dinitrosalicylic acid
dNTP	Deoxyribonucleotide

G. thermoglucosidasius	*Geobacillus thermoglucosidasius*
GH	Glycoside Hydrolase
HPLC	High-performance liquid chromatography
HUS	Hemicellulose Utilisation System
LDH	Lactate dehydrogenase
MilliQ	Ultrapure water
ORF	Open Reading Frame
PASC	Phosphoric-Acid Swollen Cellulose
p4070	Promoter of ORF peg4070
pAdhE	Acetaldehyde Dehydrogenase promoter
PCR	Polymerase Chain Reaction
PDC	Pyruvate decarboxylase
PDH	Pyruvate dehydrogenase
PFL	Pyruvate formate lyase
pRPLS	RplS constitutive promoter
pSorb	Galactitol utilisation operon promoter
pTre	Trehalose utilisation operon promoter
pUP2n38	Uracil phosphoribosyltransferase constitutive promoter
RFA	Renewable Fuels Association
Sec-SRP System	Secretion-Signal Recognition Peptide System
SLH domain	Surface-Layer Homology domain
T. fusca	*Thermobifida fusca*
T. maritima	*Thermotoga maritima*
TAE	Tris-acetate-EDTA buffer
TAT System	Twin Arginine Targetting System
***Tf*cel48A**	*T. fusca* XY reducing end exoglucanase A
***Tf*cel6B**	*T. fusca* XY non-reducing end exoglucanase 6B
***Tf*LPMO10A**	*T. fusca* XY lytic polysaccharide monooxygenase A
***Tf*LPMO10B**	*T. fusca* XY lytic polysaccharide monooxygenase B
***Tm*cel12A**	*T. maritima* DSM3109 endoglucanase A
***Tm*cel12B**	*T. maritima* DSM3109 endoglucanase B
***Tm*cel5A**	*T. maritima* DSM3109 endoglucanase 5A
U	Unit of enzyme activity (µmol/min)
USM	Modified Ammonia Salts medium
V	Volt

Chapter 1: Introduction

1.1 Demand for organic compound production from sustainable resources

The escalation of oil prices since the 1970's has meant that research into renewable energy and biofuel production has become an increasing priority of the scientific community, with the production of bioethanol (C_2H_5OH) being the most established to date (Rass-Hansen *et al.*, 2007, Helm, 2011). Bioethanol has a higher octane rating and burns more cleanly compared to conventional gasoline (Demain *et al.*, 2005) and is already blended in more than 97% of gasoline sold in the U.S. today ((RFA), 2015). Moreover, ethanol itself is easily biodegradable and has low toxicity, causing little environmental pollution (Hansen *et al.*, 2005).

To illustrate the size of the ethanol industry empirically; in 2014 alone, the United States of America produced 14.3 billion gallons of ethanol (da Silveira and Mattos, 2015). That same year, U.S. ethanol exports were valued at $2.1 billion, the second-highest on record. From a macroeconomic perspective, the ethanol production industry added $52.7 billion to the nation's Gross Domestic Product, highlighting the importance to the general economy of the production of this single molecule ((RFA), 2015) **(Figure 1)**.

However, the most promising emerging market for commercial fermentation systems is the production of fine chemicals and organic products (Reviewed in (Gallezot, 2012)). The diversity of organic products that human beings encounter and interact with, daily, may open up much greater commercial opportunities than the biofuel industry. In a way, the diversity of commercial organic compounds provides confined product niches that SMEs can exploit to fill the economic and environmental need for sustainability.

For example, Corbion, formerly Purac, has focused most of its endeavours on the production of biobased poly-lactic acid (PLA), and has recorded an increase in net sales of biobased innovations from €5.2 million to €10.2 million from 2013 to 2014 alone (Corbion, 2014). Similarly, Green Biologics have signed deals worth around $15 million with two Chinese biochemical businesses to provide their novel Acetate-Butanol-Ethanol (ABE) production process (Harvey, 2010). Ambitiously, Lanxess have focused their efforts on the production of the first bio-based ethylene-propylene-diene (EPDM) rubber, with a EPDM production plant currently under construction in Changzhou, China (Lanxess, 2014).

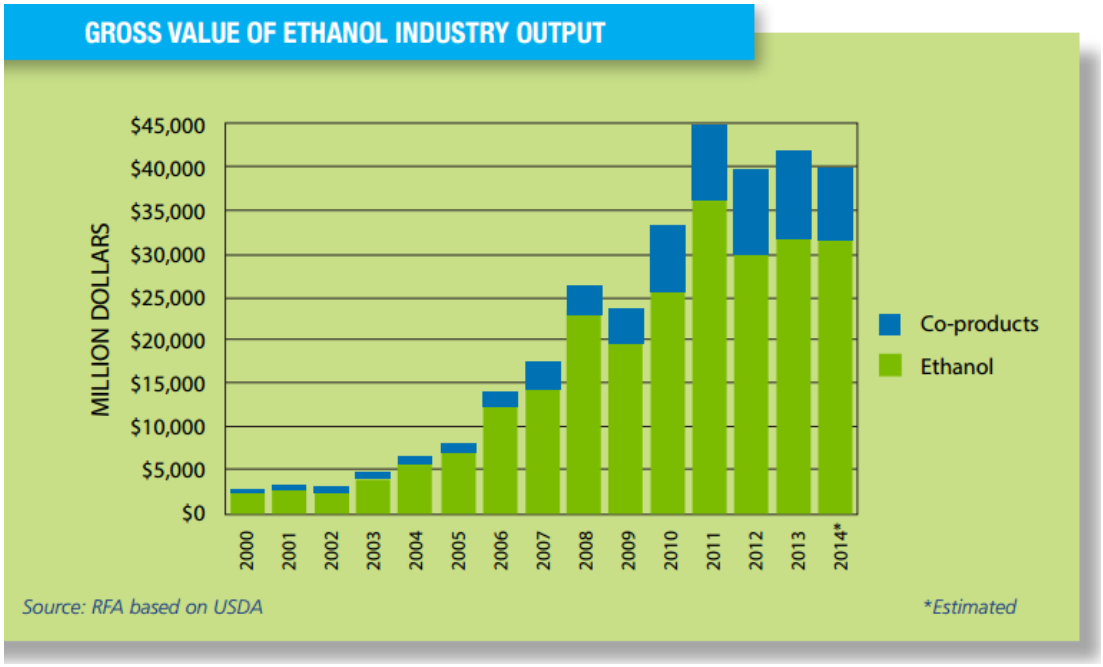


Figure 1 – Gross Annual Valuation of Ethanol Industry and Co-products Industry Output ((RFA), 2015). The dramatic increase in the value of ethanol production (green) is followed by increasing growth in the market for nutrient-dense co-products (blue), which are composed of the remaining protein, fat, and undigested fibre from the bioprocess. These make considerable contributions to the global animal feed supply as feed to beef cattle, dairy cows, swine, poultry, and fish.

Nevertheless, bioethanol remains the sustainable chemical with the highest level of commercial production to date (Limayem and Ricke, 2012). The vast majority of bio-based bioethanol is currently produced by first generation processes that involve the growth of fermentative organisms on sugars derived from sugar cane, or starch from corn and wheat (Reviewed in (Naik *et al.*, 2010)). Although these arable starches are readily hydrolysed with amylases, there is a limit to biofuel production from these sources, above which food supplies and biodiversity are threatened. With increasing global concerns over food shortages, and crop droughts becoming more and more frequent, scientific focus has shifted to second generation processes that involve the growth of fermentation organisms on sugars derived from lignocellulosic feedstocks.

1.2 Second Generation Bioprocessing Technologies

In addition to cellulose being the world's most abundant organic polymer, cellulosic biomass is the most abundant industrial waste product of modern times (Chandel and Singh, 2011). Therefore, it is no surprise that the commercial and biotechnological potential of utilising this

waste lignocellulose for the production of valuable organic products has resulted in intense research (Mazzoli *et al.*, 2012). Second generation bioprocessing circumvents the aforementioned limitations of first-generation processes by extracting sugars from more recalcitrant non-food bioenergy crops, such as switchgrass, poplar wood, and *Miscanthus*; industrial waste products, such as fruit pulp, woodchips and skins; or residual stems, husks and leaves from food crop processing (Inderwildi and King, 2009).

A non-food bioenergy crop of increasing global interest is the large, perennial grass hybrid *Miscanthus* × *giganteus*, which produces more mass overall and greater bioethanol yields than corn. For example, 12 million hectares of *Miscanthus x giganteus* can provide 133×10^9 L of ethanol, whereas corn grown from the same area of land can only provide 49×10^9 L, while requiring greater nitrogen and fossil energy inputs in its cultivation (Heaton *et al.*, 2008). Another advantage to *Miscanthus x giganteus* is the hybrid's sterility, requiring underground propagation through rhizomes, and this non-invasive quality is attractive for growth in areas foreign to *Miscanthus x giganteus* (Lewandowski *et al.*, 2000). In fact, recent research even demonstrates added benefits of *Miscanthus x giganteus* with a potential ability to sequester carbon into the earth (Clifton-Brown *et al.*, 2007).

1.3 Lignocellulosic Biomass composition

Although there is great variation in the chemical composition of lignocellulosic bioenergy crops, the general structure of lignocellulosic biomass can be divided into three components.

Cellulose

The core component of lignocellulosic biomass is crystalline cellulose, which determines the robust plant wall structure. Again, although the chemical composition of plant cell walls varies, the cellulose content usually accounts for 35–50 % of the dry weight. Cotton fibre is the only natural pure cellulose; its cellulose content reaches 95–97 % (Haigler *et al.*, 2012).

These carbon-rich cellulose molecules arrange regularly into fibrils, composed of entwined micro-fibrils, making cellulose stronger than steel wire of the same thickness (Chen, 2014). In turn, each micro-fibril is formed with elementary fibrils arranged in parallel and linked by β-1,4-glycosidic bonds.

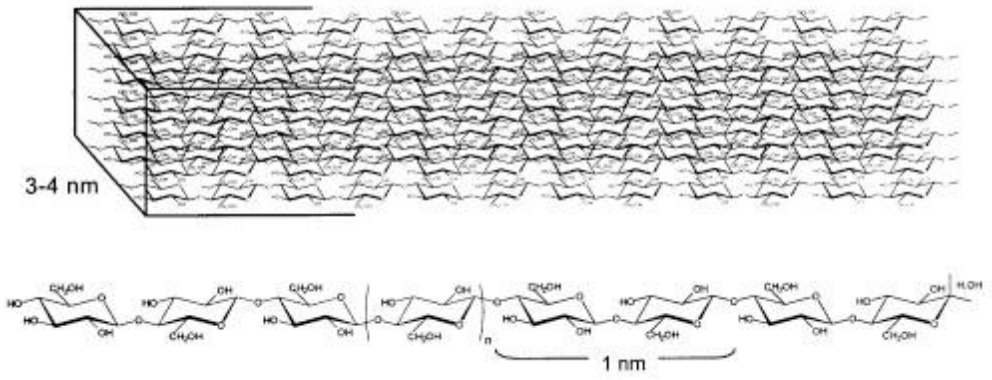


Figure 2 - Structure and association of cellulose. Image obtained from Sun, 2010. (A) The tight packing of cellulose chains in a 3-4 nm cellulose fibril. (B) Individual cellulose chain, with an indication of the length of the structural unit, cellobiose. A sub-monomer glucose is flanked by parentheses.

A key aspect of the packing of cellulose fibrils is the interplay between the crystalline and non-crystalline phases of cellulose **(Figure 2)** (Schwarz, 2001). The hydrogen bonding between the hydroxyl groups of one cellulose chain and the oxygen present in another cellulose chain can lead to the formation of very tightly-packed cellulose crystallites (Chen, 2014). These linkages result in the aforementioned structural rigidity of cellulose, but provide remarkable resistance to chemical or enzymatic hydrolysis and insolubility in common solvents (Deguchi *et al.*, 2008). In fact, crystalline cellulose is so robust that it cannot be broken up no matter how long it is cooked in boiling water, whereas starch, the geometric isomer of cellulose with α-1,4-glycosidic bonds, is readily solubilised in water at 60°C.

On the other hand, the cellulose crystallisation process is not perfect and the absence of ordered hydrogen bonding results in amorphous cellulose, which is more susceptible to degradation than the crystalline regions (Chundawat *et al.*, 2011). Addressing the recalcitrance of cellulose is at the heart of evolving technologies focused on extracting the maximum sugar yields in the efficient conversion of lignocellulosic biomass.

Hemicellulose

In contrast to cellulose, hemicellulose is a copolymer composed of different amounts of several saccharide molecules, commonly branched xylan and arabinan polysaccharides, which interact with cellulose to form a network with the microfibrils. This hemicellulose component accounts for 20-35% of lignocellulosic dry matter (Bunnell *et al.*, 2013).

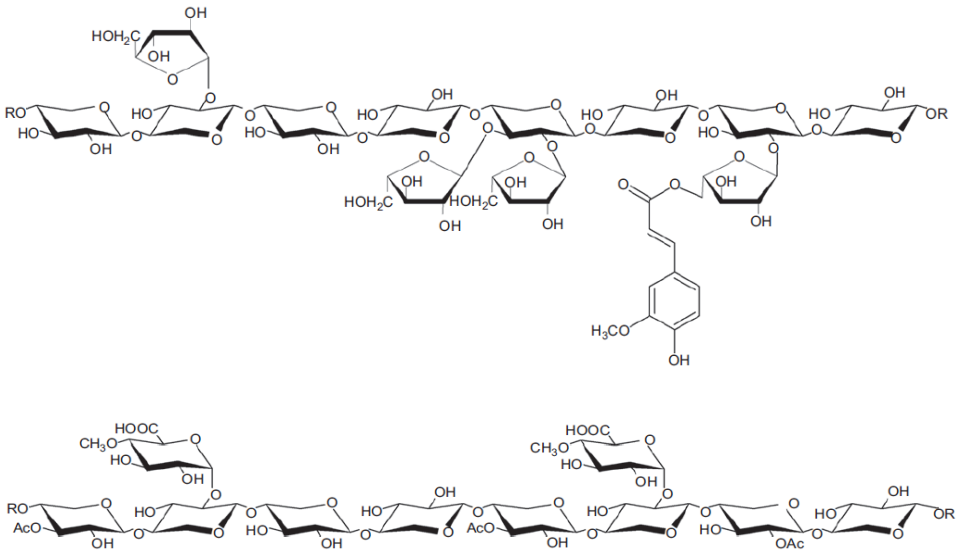


Figure 3 – The chemical structures of three hemicellulose polysaccharides. Image obtained from Pollet *et al.*, 2010. (A) Arabinoxylan is commonly found in cereal grains. (B) Glucuronorabinoxylans are commonly found in hardwoods.

Almost all plants contain xylan. The main chain of xylan is a linear homopolymer of linked D-xylosyl residues, and the diversity of hemicellulose structures in different plant groups arises from the different branches that are attached to this 1,4-β-D-xylopyranose backbone (Comino *et al.*, 2014). For example, whereas cereal grains commonly contain arabinoxylans, which are composed of a 1,4-β-D-xylopyranose backbone largely substituted with D-arabinose (Courtin and Delcour, 2002) **(Figure 3A)**, hardwoods commonly contain glucuronoxylans, which are composed of the same 1,4-β-D-xylopyranose backbone substituted with 4-O-methyl-D-glucuronic acid (Pinto *et al.*, 2005) **(Figure 3B)**.

Lignin

Further encapsulating this polysaccharide mesh is lignin, an extremely heterogenous aromatic polymer. The basic units of lignin are made up of three different phenyl propane monomers (coniferyl alcohol, syringyl alcohol and coumaryl alcohol), which are polymerised by a free radical coupling reaction. The free radical polymerisation process generates a series of random ether and carbon–carbon bonds, although the dominant linkage observed is the stable β-O-4 ether bond (Li, 2009).

Intriguingly, similarities can be drawn between the natural relationship of lignin and cellulose in providing a structural function in plants to that epoxy resin and glass fibres in a fiberglass boat. Cellulose fibres, similar to glass fibres, function as the primary load-bearing elements,

whereas lignin, similar to epoxy resin, provides additional stiffness and rigidity to the complex (LignoWorks, 2015).

Although the utilisation of lignin is not covered in this project, the large amounts of lignin produced as a by-product of lignocellulosic biomass degradation is attractive as a tertiary source of potentially utilisable carbon. In fact, several *Geobacillus* spp. have been shown to metabolise aromatic compounds (Adams and Ribbons, 1988), and putative coumaric acid utilisation operons have been identified. A process that can additionally convert a portion of the released lignin to valuable organic s may, in the future, be one avenue to improve the economic returns of lignocellulosic biomass bioprocessing.

1.4 Conventional Bioprocessing

Traditionally, the bioconversion of lignocellulosic biomass into bioethanol, or other organic compounds, involves an economically inefficient two-step process with separate compartments and conditions segregating the depolymerisation of lignocellulosic polysaccharides to fermentable sugars and the biological fermentation process **(Figure 4)** (Chundawat *et al.*, 2011).

Initially, mechanical shearing and chemical-based pre-treatments of the biomass substrate can reduce the crystallinity of the cellulose and provide greater enzyme access by either partially hydrolysing hemicellulose or lignin (Reviewed in (Alvira *et al.*, 2010)). Common pretreatment techniques include acid hydrolysis, steam explosion, ammonia fibre expansion (AFEX), alkaline wet oxidation, and hot water pretreatment (Alvira *et al.*, 2010). However, ethanol production has been reported from untreated whole biomass, including whole untreated sugarcane (Pereira *et al.*, 2015) and untreated corn stover (Lau and Dale, 2009). Nevertheless, the subsequent bioprocessing to commercial organic products remains the same regardless of this optional pre-treatment step.

In the first stage of conventional bioprocessing, biomass is treated with commercial glycosyl hydrolase cocktails to liberate the component mono-, di- and oligosaccharides (Talebnia *et al.*, 2010). These glycosyl hydrolase cocktails are often optimised for a specific lignocellulosic biomass substrate, and their production is a lucrative business. For example, Novozymes, the world's largest enzyme manufacturer, attributes 18% of its $1.6 billion turnover to selling enzymes to the first-generation biofuel industry (Slade, 2010). However, the price of these cocktails is one of the main contributors to the costs of conventional biofuel production.

The product of this first hydrolysis step is a sugar-rich mixture that is then fed to a fermentative microorganism for intracellular bioprocessing to the biofuel or desired organic product (Demirbas, 2009). Bioethanol-producing microorganisms are commonly bacterial (Romero *et al.*, 2007, Ingram *et al.*, 1987, Cripps *et al.*, 2009) or fungal (Lee *et al.*, 2008, Alper *et al.*, 2006, Martinez *et al.*, 2008) strains that have been metabolically engineered for optimal utilisation of pentose and hexose monosaccharides, and subsequent bioconversion of pyruvate to bioethanol.

Although conventional bioprocessing is an established regime for 2G production of biofuels and organic products, it still requires economic improvement (Juneja *et al.*, 2013). The cost of the cellulase cocktails must be factored into the process costs. This provides challenge for evaluation of alternative bioprocessing schemes that can potentially circumvent the need, or at least minimise the requirement for expensive cellulase cocktails.

1.5 Consolidated Bioprocessing

A recent concept in lignocellulose bioprocessing is Consolidated Bioprocessing (CBP), wherein the processes of polysaccharide hydrolysis and metabolite production are combined into a single step (Reviewed in (Lynd *et al.*, 2005)). A CBP organism is therefore a single biocatalyst for the direct conversion of pre-treated lignocellulosic material into a specific metabolic product **(Figure 5)**. However, a natural CBP organism, capable of direct conversion of polysaccharides to a desired metabolic product in a single bioreactor, has yet to be reported.

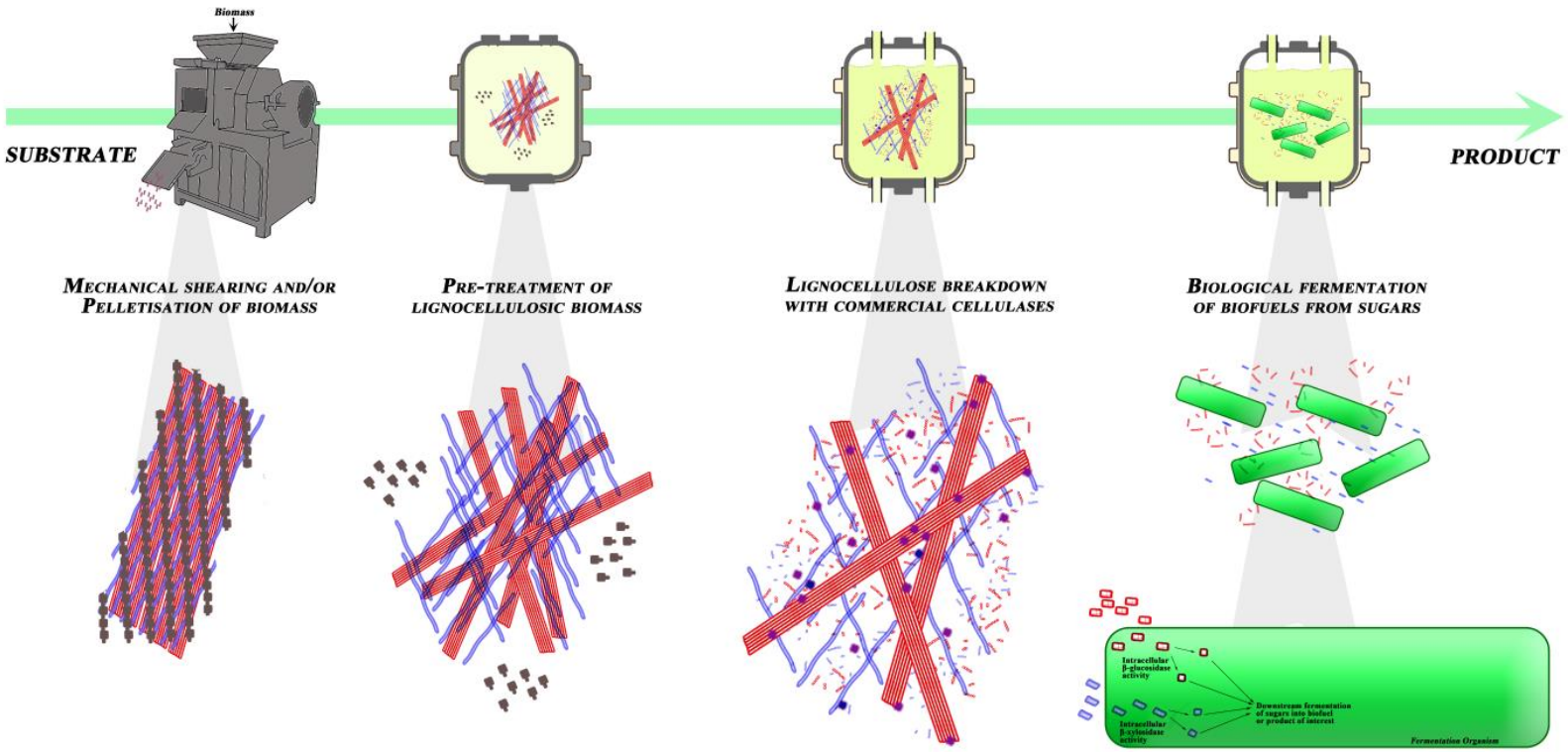


Figure 4 – Conventional Bioprocessing Technologies. The process for both first generation and second generation conventional bioprocessing often involves the initial pelletisation of lignocellulosic biomass to increase the available surface area for hydrolysis. A subsequent pre-treatment step serves to open up the macromolecular structure of lignocellulosic biomass. For example, 20% ammonia soaking has been shown to hydrolyse lignin into soluble components (brown hexagons), and stretch out the structures of cellulose (red fibres) and hemicellulose polysaccharides (blue fibres) (Bartosiak-Jentys, unpublished). This improves the hydrolytic rates of glycosyl hydrolase cocktails, which are typically a mixture of cellulases (red circles) and hemicellulases (blue circles) for the synergistic digestion of both cellulose and hemicellulose fractions. Liberated mono-, di- and potentially oligosaccharides are fed in the subsequent steps to engineered fermentation organisms (green), which internalise and channel the sugars to the organic product of choice.

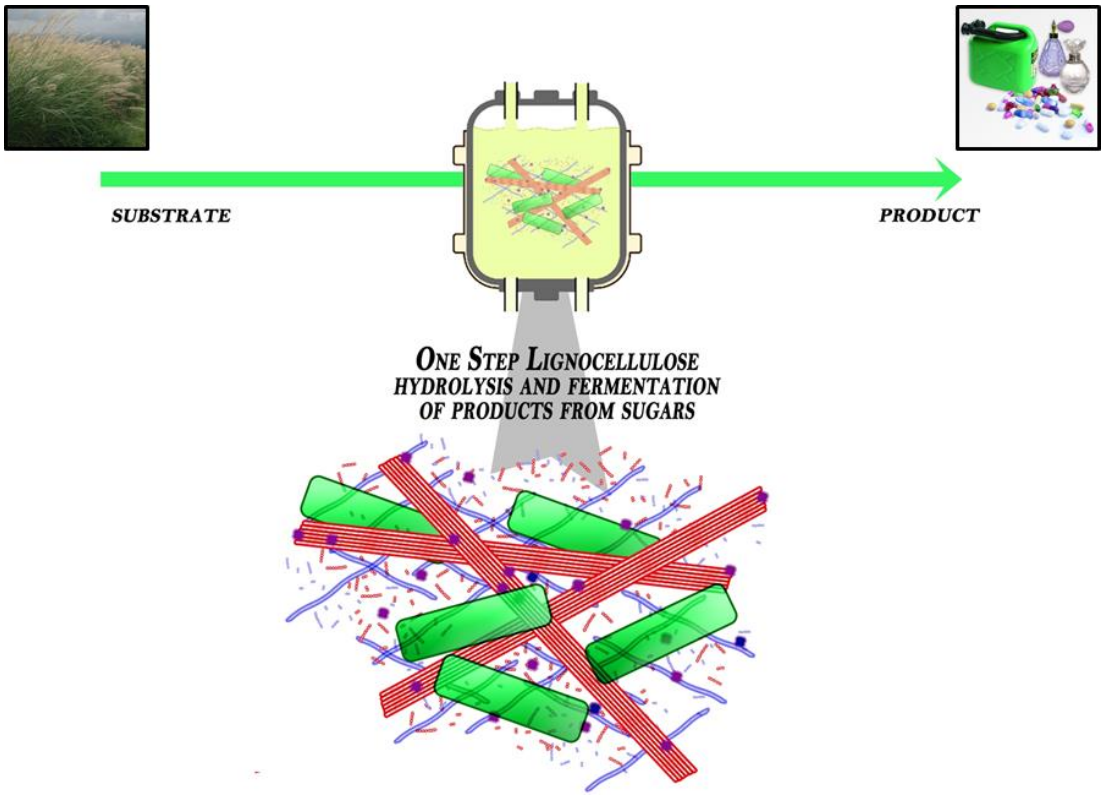


Figure 5 – Consolidated Bioprocessing Technologies. From substrate to product, the two processes of polysaccharide breakdown and biological fermentation are merged into a one-step Consolidated Bioprocess. Here, the organism has encoded capabilities to express and secrete the cellulases (purple circles) and hemicellulases (blue circles) required to breakdown the cellulose (red fibres) and hemicellulose polysaccharides (blue fibres), and simultaneously consumes the liberated sugars in order to produce the organic compound of choice.

Efforts to engineer CBP microbial strains can be divided into two strategies. CBP Strategy 1 involves the metabolic engineering of naturally cellulolytic microorganisms to improve their fermentative capabilities and efficiencies (Wood and Ingram, 1992, Jin *et al.*, 2011). Conversely, CBP Strategy 2 involves the genetic engineering of efficient commercial fermentation microorganisms to express a heterologous cellullolytic system and facilitate the pre-fermentation stage *in vivo* (la Grange *et al.*, 2010, Deshpande, 1992, Jin *et al.*, 2012, Patel *et al.*, 2005, Zhang *et al.*, 2009, Ou *et al.*, 2009).

Attempts to engineer Strategy 1 CBP organisms have produced impressive results of late. In 2014, an *ldh− adhE+* strain of the cellulolytic *Caldicellulosiruptor bescii* was demonstrated to produce ethanol from unprocessed switchgrass, an abundant and economically sustainable lignocellulosic plant biomass. Intriguingly, ethanol made up 70% of

detected fermentation products during growth on cellobiose, Avicel, and switchgrass (Chung *et al.*, 2014). The Liao group have engineered isobutanol pathway genes into the cellulolytic *C. thermocellum*, resulting in a strain reported to produce 5.4 g/L of isobutanol from cellulose in minimal medium within 75 h (41% of theoretical yield) (Lin *et al.*, 2015).

Reported attempts to engineer Strategy 2 CBP organisms have also produced promising results. The Keasling group at the University of California have already published the engineering of *Escherichia coli* (*E. coli)* strains capable of direct productionof fatty-acid ethyl esters, butanol, and pinene from pretreated switchgrass (Bokinsky *et al.*, 2011). The same group reported the upregulation of an endogenous endoglucanase in *B. subtilis* to facilitate the production of greater yields of lactic acid directly from amorphous cellulose and some types of pretreated biomass without addition of organic nutrients (Zhang *et al.*, 2011).

As expected in such an emerging field of research, there is growing disparity over the perquisites that constitute a CBP organism. For example, Yanase *et al* have reported the expression of a *R. albus* β-glucosidase in *Zymobacter palmae* and *Zymomonas mobilis* strains to confer the capability to produce ethanol from cellobiose at more than 95% of the theoretical yield (Yanase *et al.*, 2005). In contrast, Vasan *et al* reported the expression of an *E. cloacae* endoglucanase in *Z. mobilis* to facilitate ethanol fermentation at 5.5% v/v from CMC and 4% v/v from NaOH-pretreated bagasse (Vasan *et al.*, 2011). It can be argued that the latter organism's ability to grow on a complex polysaccharide and a pre-treated biomass is an excellent example of an engineered CBP organism, while the first example does not really constitute a CBP organism with the sole ability to grow and ferment from a disaccharide.

The definition of a CBP organism that shapes the objectives of this project is a fermentative organism that has the capability to depolymerise complex polysaccharides present in lignocellulosic feedstock into their constituent oligo-, di- and monosaccharide components, followed by the transmembrane transportation and fermentation of these liberated sugars into the desired organic fermentation product. The limited success to date in CBP engineering can be attributed to three issues. Firstly, the engineering of fermentative capabilities in a cellulolytic organism requires considerable metabolic engineering (and therefore development of genetic engineering tools) and strain development to optimise production yields. Secondly, the expression of cellulolytic capabilities in mesophiles is limited by the general thermophilic nature of cellulolysis. That is, the cellulolytic enzymes with the

greatest efficacy on lignocellulose often perform best at elevated temperatures (50-80°C), which are above the growth temperatures of mesophilic fermentation organisms. In fact, studies that have compared cellulolytic systems from thermophiles and mesophiles have shown that thermophilic systems tend to outperform their mesophilic counterparts at elevated temperatures (Mingardon *et al.*, 2011, Viikari *et al.*, 2007). Lastly, for optimal utilisation of the products of lignocellulosic biomass degradation, the CBP organism must have the catabolic versatility to ferment both pentose sugars, derived from hemicellulose breakdown, and hexose sugars, derived from the hemicellulose and cellulose fraction. This last issue has been a major drawback with *S. cerevisiae* and other commercial fermentation organisms (Ha *et al.*, 2011). However, the increasing characterisation of pentose-fermenting yeasts may provide promising alternatives to conventional yeasts (Reviewed in (Hahn-Hagerdal *et al.*, 2007)).

1.6 *Geobacillus thermoglucosidasius*

A potential host for the development of CBP is the thermophilic, facultatively-anaerobic *Geobacillus thermoglucosidasius*, which is capable of growth between 50°C and 70°C (Thompson *et al.*, 2008) and employs a mixed acid fermentation pathway. The natural ability of this organism to ferment major hexose and pentose sugars (Thompson *et al.*, 2008) at a high temperature has generated significant interest in its industrial potential. As a result, the genomes for the C56-YS93, TNO-09.020 and Y41.MC1 strains have been sequenced, annotated and made publically available (NCBI Accession - CPO02835, CM001483 and CP002293 respectively). The genome of the ethanol-producing industrial TM242 strain has also been sequenced and annotated as a commercial venture funded by TMO Renewables Ltd. An added advantage in the use of *G. thermoglucosidasius* is its relatively-close phylogenetic relationship to the Gram-positive model organism *Bacillus subtilis*, which provides a benchmark for improving the *Geobacillus* genetic toolkit and furthering biotechnological applications (Nazina *et al.*, 2001).

The catabolic versatility of *Geobacillus* spp., and their ability to secrete commercially useful enzymes, such as hemicellulases and amylases, are currently being exploited for both biocatalysis and metabolic engineering. Moreover, there a number of other diverse applications also being investigated, from production of sweeteners, production of therapeutics and the exploitation of *Geobacilus* spp. S-layer-based nanostructures (Reviewed in (Hussein *et al.*, 2015)).

Nevertheless, the most widely researched biotechnological application of *G.thermoglucosidasius* is for fermentation to produce second generation biofuels, an endeavour largely arising from the industrial impetus provided by companies such as Agrol Ltd and TMO Renewables Ltd. Growth at 60-70^{o}C facilitates continuous removal of volatile fermentation products (e.g. the boiling point of ethanol is 78°C) while not causing excessive attrition of mechanical equipment (Cripps *et al.*, 2009). Furthermore, higher growth temperatures reduce potential contamination issues from mesophilic contaminants (Cripps *et al.*, 2009). However, it is the catabolic promiscuity of *G. thermoglucosidasius*, particularly its ability to take up and degrade a wide range of oligomeric carbohydrates, that sets it apart in terms of second-generation bioprocess design (Cripps *et al.*, 2009).

Ethanol is a natural, but secondary, fermentation product of *G. thermoglucosidasius*. After knocking out the L-lactate dehydrogenase pathway, it was expected that the fermentation products would be determined by the residual pyruvate-formate-lyase (PFL) pathway. However, higher yields of ethanol were obtained than expected from the PFL pathway and it was recognised that, as in *B. subtilis*, pyruvate dehydrogenase (Pdh) was still active under anaerobic conditions. By knockout out the PFL pathway and upregulating expression of Pdh a novel, redox balanced homo-ethanol pathway was developed, which forms the basis of an industrial process (Cripps *et al.* 2009). Ethanol yields from glucose of greater than 90% of the theoretical value have been achieved in the triple mutant (Δ*ldh*, Δ*pflB, pdh*$_{up}$) process strain *G. thermoglucosidasius* TM242 (Cripps *et al.*, 2009) with productivities as high as 2.85 g/L/h on glucose and 3.2 g/L/h on cellobiose (Cripps *et al.*, 2009).

In fact, the ethanologenic capabilities of *G. thermoglucosidasius* TM242 may further expand in the future with the potential introduction of alternative pathways for ethanol fermentation. One potential avenue for increasing flux to ethanol is via pyruvate decarboxylase (PDC, EC 4.1.1.1), the fermentation route used by *Saccharomyces cerevisiae* and the bacteria *Zymomonas mobilis* and *Zymobacter palmae*. PDC catalyses the non-oxidative decarboxylation of pyruvate to acetaldehyde which is then converted to ethanol by alcohol dehydrogenase (ADH, EC 1.1.1.1). Although PDCs from plants and yeast have been widely studied, only a few bacterial examples have been identified and characterised, notably from *Zymomonas mobilis, Zymobacter palmae, Acetobacter pasteurianus,* and *Sarcina ventriculi* (Raj *et al.* 2002). So far, a PDC of thermophilic origin has not be discovered, and heterologous expression of both *Z. mobilis* and *Z. palmae* PDC in *G. thermoglucosidasius* does

not result in functional enzyme activity at temperatures exceeding 55 °C (Thompson *et al.*, 2008). However, both enzymes were reported to show good *in vitro* thermostability at these temperatures, and previous studies indicate the native (prefolded) *Z. mobilis* PDC being stable up to 60°C (Thompson *et al.*, 2008, Pohl *et al.*, 1995). Recently, a PDC from *Gluconobacter oxydans* which remains thermostable *in vitro* at 45°C was expressed in *G. thermoglucosidasius* TM89, a Δ*ldh* variant of the NCIMB 11955 strain and grown fermentatively at 52°C, resulting in yields as high as 0.35 ± 0.04 g/g ethanol per gram of glucose consumed (Thompson *et al.*, 2008).

Nevertheless, the catabolic versatility and ethanologenic capability of *G thermoglucosidasius* make the organism a strong candidate for CBP engineering. That is to say, *G. thermoglucosidasius* already has the natural capability to utilise most of the hydrolysis products of lignocellulosic biomass degradation and subsequently ferment them into desired organic products, with bioethanol the current product of choice. Moreover, the elevated temperatures (60-70°C) that *G. thermoglucosidasius* grows at aligns well with the optimum temperatures of the most effective cellulolytic enzymes characterised to date.

1.7 Natural degradation of lignocellulosic biomass

G. thermoglucosidasius strains sequenced to date do not produce any cellulose-degrading enzymes, although they do grow rapidly on cellobiose. Surprisingly, given the evidence for growth of xylose oligomers, the ethanologenic *G. thermoglucosidasius* TM242 mutant that will be the chassis for this project also does not encode any secreted hemicellulases either. Since the requirement for heterologous expression and secretion of cellulases and hemicellulases is a fundamental component of the CBP project strategy, a wider introduction to these enzyme classes is useful.

The hydrolysis of the amorphous and crystalline cellulose into its constituent glucose monosaccharides is accomplished by the catalytic activities of cellulases, a subfamily of the glycosyl hydrolase (EC 3.2.1.-) enzyme group. Cellulases are further subdivided according to the hydrolysis reaction they catalyse (Lynd *et al.*, 2002).

Exoglucanases

Exoglucanases act on either the reducing or non-reducing ends of crystalline cellulose to liberate the disaccharide cellobiose (Birsan *et al.*, 1998). Exoglucanases are characterised by their effective activity on crystalline cellulose, the major component of lignocellulosic biomass, and thus are regarded as the workhorses of lignocellulose degradation. This is exemplified by the high exoglucanase expression levels exhibited by both *T. fusca* and *C. thermocellum* (Kruus et al., 1995b, Irwin et al., 2000). Notwithstanding, these high exoglucanase expression levels are likely to have evolved to compensate for their characteristically low specific activities on recalcitrant crystalline cellulose (Calza *et al.*, 1985).

Nevertheless, reducing-end acting exoglucanases, belonging to the glycosyl hydrolase 48 family, and non-reducing-end acting exoglucanases, belonging to the glycosyl hydrolase 6 and 9 families, constitute the major components of all effective crystalline cellulose degradation systems.

Endoglucanases

Endoglucanases generate shorter oligosaccharides by hydrolysing internal sites in amorphous cellulose and longer oligosaccharide chains (Reviewed in (Levy *et al.*, 2002)). In contrast to exoglucanases, endoglucanases are far more diverse in nature, widely-reported, observed in a multitude of glycosyl hydrolase families and exhibit considerably higher specific activities on amorphous cellulose, their preferred polysaccharide substrate (Levy *et al.*, 2002). Although cellulolytic *C. thermocellum* encodes four exoglucanases, it encodes over 14 endoglucanases, which are differentially-expressed based on the lignocellulose substrate the organism is grown on (Wei *et al.*, 2014). This observation illustrates the importance of secreting optimal endoglucanases, with subtle differences in substrate specificities, for effective utilisation of polysaccharides by cellulolytic organisms. To put it another way, whereas the decrystallising-activities of exoglucanases are the rate-limiting step of effective lignocellulose breakdown, the hydrolytic activities of endoglucanases play a similarly major role in the fast release of sugars from the released amorphous cellulose.

Table 1 – Exoglucanase- and, endoglucanase-, and xylanase-containing glycosyl hydrolases (GH) families. Generally, the hydrolysis of glycosidic bonds is catalysed by a general acid proton donor and a nucleophile, which can be either D (Aspartic Acid) or E (Glutamic Acid) (Davies and Henrissat, 1995). The characterised stereochemical outcome of the reaction is italicised.

GH family	Enzymatic function	Nucleophile/ Proton donor
GH1	*Retaining /* β-glucosidase; β-galactosidase; β-mannosidase; β-glucuronidase	E/E
GH2	*Retaining /* β-galactosidase; β-mannosidase; β-glucuronidase	E/E
GH5	*Retaining /* endoglucanase; cellobiohydrolase (non-reducing end); endo-1,4-xylanase	E/E
GH6	*Inverting /* endoglucanase; cellobiohydrolase (non-reducing end)	D/D
GH7	*Retaining / endoglucanase;* cellobiohydrolase (*reducing-end*)*; chitosanase*	E/E
GH8	*Inverting /* chitosanase; endoglucanase; licheninase; endo-1,4-xylanase	D/E
GH9	*Inverting /* endoglucanase; cellobiohydrolase (*non-reducing end*); β-glucosidase; β-glucosaminidase	D/E
GH10	*Retaining /* endo-1,4-xylanase; endo-1,3-xylanase	E/E
GH11	*Inverting /* endo-1,4-xylanase; endo-1,3-xylanase	E/E
GH48	*Inverting /* cellobiohydrolase (*reducing end*); endoglucanase; chitinase	?/E

β -Glucosidases

β-glucosidases hydrolyse the disaccharide cellobiose to glucose (Withers, 2001). *Geobacillus* spp. generally encode an intracellular β-glucosidase to facilitate intracellular hydrolysis of cellobiose to glucose (De Maayer *et al.*, 2014). This serves two advantages. Firstly, in the competitive soil flora ecosystem where *Geobacillus* spp. are generally found, the uptake of cellobiose provides an important competitive advantage against non-cellobiose utilisers for survival in conditions of carbon limitation. Secondly, uptake of a cellobiose molecule or a glucose molecule by either the phosphotransferase system (PTS system) or ATP-Binding

Cassette transporters (ABC transporters) both require the use of one phosphoenolpyruvate (PEP) molecule or ATP molecule, respectively. Therefore, the uptake of cellobiose reduces the metabolic burden of sugar uptake by capturing two glucose molecules at the cost of a single ATP.

In fact, the sugar uptake capabilities of *G. thermoglucosidasius* are thought to extend to cellulo-oligosaccharides as long as 5 glucose monomers in length (cellopentaose) (S. Martin, ZuvaSyntha Ltd., personal communication). Therefore, there is no requirement for the heterologous secretion of β-glucosidiases in the envisioned CBP system. Rather, the present strategy is the reduction of oligosaccharides to a sufficient length for efficient uptake.

Carbohydrate-Binding Modules (CBMs)

A subset of these cellulases and other members of the glycosyl hydrolase family contain discrete modules with carbohydrate-binding activity. These folds, termed Carbohydrate-Binding Modules (CBMs), are involved in the recognition of both of amorphous and crystalline forms of cellulose (Jamal *et al.*, 2004) and even promote the enzymatic deconstruction of intact plant cell walls (Hervé *et al.*, 2010). Therefore, CBM-containing glycosyl hydrolases are a valuable asset for the degradation of lignocellulosic biomass in an engineered cellulolytic system.

Hemicellulases

Although the hemicellulosic component of bioenergy crops is less abundant than the cellulosic component (Kumar *et al.*, 2008), it may provide an important source of readily fermentable hexose and pentose sugars due to the considerably lower recalcitrance of hemicellulose in comparison to crystalline cellulose. Due to the variation in composition, branching, and linkages between the different hemicelluloses, there is remarkable diversity between the different hemicellulase glycosyl hydrolase families with regards to their structure and activity (Birsan *et al.*, 1998).

Since the major hemicellulose found in bioenergy crops is xylan, the emphasis of this project will focus on xylanases, which are commonly found in glycosyl hydrolase 10 and 11 families (Kulkarni *et al.*, 1999).

Cell surface attached cellulases

The attachment of proteins to the bacterial cell surface through the function of cell wall binding domains (CWBDs) has been widely-reported.

Surface-Layer Homology (SLH) domains are a major example of CWBDs, and have been observed on the amino- or carboxyl-termini of many structural proteins and enzymes (Schneewind and Missiakas, 2012), including glycosyl hydrolases (Kosugi *et al.*, 2002). These domains are found in double- or triple-repeats that non-covalently attach to either teichoic acid or teichuronic acid in the bacterial cell wall. In Gram-positive bacteria, binding of proteins through SLH domains forms a hexagonal-array of proteins commonly covering the cell surface, known as the Surface-Layer or S-Layer (Mesnage *et al.*, 2000). This S-Layer is thought to facilitate a number of functions, including pH-resistance, temperature-resistance and antigen-display (Schuster and Sleytr, 2014). By attaching SLH domains to the C-terminus of the *B. subtilis* 168 levansucrase, a fully functional enzyme has been successfully attached to the outer cell surface of *B. anthracis* (Mesnage *et al.*, 1999), suggesting that it is a useful anchoring mechanism. As the predicted *G. thermoglucosidasius* proteome contains proteins with SLH domains, these may be similarly exploited in attaching glycosyl hydrolases to the cell surface.

Multicomponent cellulosome systems

A well-studied example of a SLH-displaying cellulolytic bacterium is the anaerobic Gram-positive thermophile *Clostridium thermocellum*, which expresses a well-characterised multi-enzyme complex termed a cellulosome (Doi and Kosugi, 2004). These cellulosomes are composed of a large, non-catalytic scaffoldin protein that facilitates the binding of a multitude of different glycosyl hydrolases and other enzymes that contain corresponding dockerin modules, to intermittent cohesion domains along the scaffoldin structure (Bayer *et al.*, 2008).

The attachment of a multitude of cellulases in close proximity to one another increases the synergistic cellulolytic activity between them, resulting in the ability of *C. thermocellum* to grow on recalcitrant cellulosic substrates, including Avicel, filter paper, and pretreated mixed hardwood (Bayer *et al.*, 2008).

This protein multiplex is attached to the cell membrane indirectly through an anchoring protein. This anchoring protein is attached to the scaffoldin protein by Type II cohesion-dockerin interactions and is strongly attached to the cell surface using Surface-Layer Homology (SLH) domains described above (Leibovitz *et al.*, 1997). The attachment of this large multi-protein complex in close proximity to the cell ensures that the majority of the liberated sugars from cellulose are absorbed and utilised by the cellulosome-expressing bacterium (Leibovitz *et al.*, 1997).

Recently, there has been progress in the heterologous expression and display of an attached cellulolytic system, referred to as mini-cellulosomes, in non-cellulolytic microorganisms. One example of these attached-CBP organisms is a genetically engineered *E. coli* LY01 strain expressing and displaying an endoglucanase, exoglucanase, and β-glucosidase from *Clostridium cellulolyticum* on its cell surface (Ryu and Karim, 2011). Coupled with the expression of ethanol-producing enzymes, this strain demonstrated ethanol production at 3.59 ± 0.15 g/L and 0.71 ± 0.12 g/L directly from phosphoric acid-swollen cellulose (PASC) and pretreated corn stover respectively.

1.8 Current Genetic Engineering tools in *G. thermoglucosidasius*

For effective CBP engineering, which fundamentally involves the heterologous expression and secretion of the aforementioned cellulases and hemicellulases, there is a need for robust and straightforward genetic manipulation of the *G. thermoglucosidasius* genomes.

The promise of *G. thermoglucosidasius* as genetically-malleable chassis for biotechnological applications is augmented by their relatively close phylogenetic relationship to the workhorse *B. subtilis*. However, the thermophilic nature of *Geobacillus* spp has meant that genetic toolkits used in *Bacillus* spp and other mesophiles are limited due to thermal instability of proteins and commonly used antibiotics. Thus, tools such as the *Lactococcus lactis* group II intron-based Targetron technology, which have been adapted for use in *C. acetobutylicum*, are incompatible for use in *Geobacillus* spp. Therefore, the development of novel thermoactive tools for the genetic engineering of *Geobacillus* spp, and other thermophilic bacteria, requires the exploitation of native genetic machinery.

Plasmid vectors

The initial developments in the genetic manipulation of *Geobacillus* spp were the characterisation and development of plasmids capable of self-replication and selection markers for plasmid maintenance through multiple generations **(Table 2)**. Plasmids that replicate via the rolling circle (RC) mechanisms and theta-replicating mechanisms have been described (reviewed by (del Solar *et al.*, 1998)). Initially, multiple vectors based on different replicons were constructed, but with drawbacks that made them inconvenient for use as genetically-malleable shuttle vectors. The *G. stearothermophilus* shuttle vector pBST22 (derived from the natural *G. stearothermophilus* plasmid pBST1) lacked a multiple cloning site and the facility for β-galactosidase-mediated blue–white screening in *Escherichia coli*, and pNW33N (derived from a *B. coagulans* cryptic plasmid pBC1) is maintained in *Geobacillus* spp using chloramphenicol, which is only moderately thermostable (Taylor *et al.*, 2008).

To improve versatility, pUCG18 was constructed by introducing the evolved kanamycin resistance gene and origin of replication (theta) from pBST22 with pUC18, retaining all of the cloning and selection benefits of the latter (Taylor *et al.*, 2008). This has subsequently been further improved as pUCG3.8 by a reduction of size (Bartosiak-Jentys *et al.*, 2013) and more recently converted into a modular format which allows ready replacement of parts, such as origins of replication and antibiotic resistance genes (Reeve et al, in preparation).

It has been argued that rolling-circle plasmids usually have a broader host range and sometimes higher plasmid copy number than their theta-replicating alternatives (Heinl *et al.*, 2011), leading to efforts to isolate, sequence and characterise new rolling-circle replicating plasmids from *Geobacillus* spp (Kananavičiūtė *et al.*, 2014).

Table 2 : *E. coli – Geobacillus* shuttle vectors. Rep refers to the characterised mechanism of plasmid replication in *Geobacillus* spp. RC = Rolling Circle. Θ = Theta Replication. Kan^R = Kanamycin resistance. Cam^R = Chloramphenicol resistance

Year of Publication	Plasmid Name	Size	Selective Marker	Origins of Replicon	Rep	Reference
2013	pUCG3.8	3.8	Kan^R *(TK101)*	pBST1	Θ	(Bartosiak-Jentys *et al.*, 2013)
2009	pTMO31	5.1	Kan^R (pUB110)	pUB110	RC	(Cripps *et al.*, 2009)
2008	pUCG18[a]	6.3	Kan^R *(TK101)*	pBST1	Θ	(Taylor *et al.*, 2008)

2001	pNW33N	3.9	Cam^R (pC194)	pBC1	RC	(Zeigler, 2001)
2001	pBST22	7.6	Kan^R *(TK101),* Cam^R (pC194)	pBST1	Θ	(Liao and Kanikula, 1990)
1993	pSTE33 [a]	5.7	Kan^R *(TK101)*	pSTK1	Θ	(Narumi *et al.*, 1993)
1992	pSTE12	5.8	Tet^R (pTHT15)	pTHT15	N/A	(Narumi *et al.*, 1992)

[a] Conjugation-mediated transfer has been reported with derivatives of these shuttle vectors incorporating an *incP* origin of transfer.

DNA Transfer

Although several procedures to transfer plasmid DNA into *Geobacillus* spp have been developed, electroporation protocols developed during the early 1990's remain the most commonly used (Narumi *et al.*, 1992, Narumi *et al.*, 1993). Frequencies of up to 2.8×10^6 transformants per µg of pSTE33 DNA were published for the isolated *G. denitrificans* K1041 (originally classified as *B. stearothermophilus*), and are the highest transformation frequencies among a large collection of *Geobacillus* strains (Zeigler, 2001). For instance, the transformation efficiencies observed by electroporation of pUCG18 DNA in *G. thermoglucosidasius* DL44 were over two orders of magnitude lower at 9.8×10^3 (Taylor *et al.*, 2008). Still, where high transformation frequency is not critical, electroporation procedures remain the preferred method for transferring DNA to *G. thermoglucosidasius*, not only due to the relative ease of the procedure, but for the facility for long term storage of electrocompetent cell preparations.

Recent demonstrations of efficient plasmid transfer into *G. kaustophilus* HTA426 (Suzuki *et al.*, 2013a, Hirokazu, 2012) and *G. thermoglucosidasius* (A. Pudney, personal communication) using conjugative transfer look set to establish an even simpler method for the routine transformation of Geobacilli. Conjugative transfer is typically performed by incubating mixtures of recipient *Geobacillus* spp cells and donor *E. coli* cells harbouring mobilisation genes found on the chromosome (*E. coli* S-17) or on helper plasmids (pRK2013 and pUB307). Exploiting their inherent thermophilicity, the recipient *Geobacillus* spp are readily distinguished from donor cells after incubation at 60°C, and conjugative transfer has been

reported to result in transfer efficiencies as high as 1.2×10^{-3} and 2.83×10^{-4} transformants per recipient *G. kaustophilus* and *G. thermoglucosidasius,* respectively.

Positive-selection markers

All endeavours in genetic engineering require indication of successful delivery of foreign DNA into the recipient strain, and selection for maintenance of the genetic construct through subsequent generations. In the research laboratory this is conventionally done using antibiotic resistance genes that confer resistance to supplemented growth inhibitors. The thermophilic nature of *Geobacillus* spp limits the use of established selection markers, with few antibiotic resistance proteins or antibiotics currently available with sufficient thermostability at 60-70^{o}C.

Of the commonly used antibiotics kanamycin has the highest thermostability, so Liao and colleagues selected (an early example of forced evolution) a thermostable variant of the kanamycin nucleotidyltransferase gene (KNT-ase), conferring resistance to the bacteriocidal antibiotic kanamycin at temperatures up to 70°C (Liao and Kanikula, 1990, Liao *et al.*, 1986). Using the *E. coli* mutD5 mutator strain to introduce mutations and selection in *G. stearothermophilus*, a thermostable KNT-ase TK101 mutant (D80Y, T130K) of the mesostable KNT-ase gene from pUB110 was developed, and has been shown to function as a selection marker in both *Geobacillus* spp. and *E. coli* (Taylor *et al.*, 2008, Bartosiak-Jentys *et al.*, 2013).

Reporter genes

Fundamental physiological studies and biotechnological applications involving single or multiple gene expression depend on investigation and application of promoter and ribosome binding site operation and strength. Recently, there has been a rapid increase in the characterisation of promoters, particularly inducible promoters that may be used for conditional expression and easily assayable reporter genes are useful tools for characterisation of these promoters. GFP (green fluorescent protein) is a commonly used reporter in mesophiles and a useful thermostable variant, superfolder GFP (sfGFP) (Pédelacq *et al.*, 2006) has been shown to work in *Thermus* spp and *Geobacillus* spp (Blanchard *et al.*, 2014) where is has been used for the assessment of various promoters. However, like the majority of fluorescent proteins, the maturation of the fluorescent chromophore requires molecular oxygen, therefore sfGFP cannot be used in oxygen deprived environments (eg under anaerobic conditions).

An alternative transcriptional reporter gene that can be used to circumvent this drawback is the *pheB* gene from *G. stearothermophilus* DSM 6285, which encodes a thermostable catechol 2,3-dioxygenase. The expression of *pheB* in the presence of 100 mM catechol results in the formation of the yellow-coloured 2-hydroxymuconic semialdehyde, which can be detected at an absorbance of 375 nm (Bartosiak-Jentys *et al.*, 2012). Enzymes associated with carbohydrate metabolism have also been exploited as expression reporters in *Geobacillus* spp, including α-amylase, β-galactosidase and α-galactosidase (Lin *et al.*, 2014, Blanchard *et al.*, 2014, Suzuki *et al.*, 2012).

1.9 Recombinant gene expression

For an effective CBP system, the desired glycosyl hydrolases must be heterologously expressed from transcriptional promoters of sufficient strength and regulation. In a broader context, there is a general demand for *Geobacillus* spp. compatible inducible promoters for the controlled heterologous expression of "toxic" proteins, which may require strong transcriptional silencing to facilitate sufficient growth of the host organism. However, a limited number of constitutive or inducible promoters have been characterised as suitable for controlled heterologous protein expression in *Geobacillus* spp.

For strong and constitutive expression in *G. kaustophilus* HTA426, the native promoter P_{sigA}, found immediately upstream of two housekeeping genes (Suzuki *et al.*, 2012), has been characterised in β-galactosidase assays. The promoter for ribonuclease H III, P_{RHIII}, isolated from *G. stearothermophilus* NUB3621, has -10 and -35 regions closely matching the consensus and has been used for constitutive expression of the fluorescent sfGFP reporter (Blanchard *et al.*, 2014). However, a drawback of strong and constitutive heterologous protein expression is the potential metabolic burden that may occur due to the diversion of metabolic resources towards heterologous protein expression, which would otherwise be available for increased fermentation yield.

One strategy for reducing the metabolic burden of heterologous glycosyl hydrolase expression on fermentation yield is to utilise promoters that are active through direct induction by the components of hydrolysed feedstocks (e.g. *Miscanthus*). A wide range of ligand-inducible promoters functional in *Geobacillus* spp. have been characterised, especially carbohydrate-inducible promoters. However, although several positively regulated promoters have been shown to facilitate controlled protein expression, many of these

promoters are functional under various other conditions (Suzuki *et al.*, 2013b, Bartosiak-Jentys *et al.*, 2013).

The *G. kaustophilus* HTA426 promoters P_{gk704}, P_{gk1859}, P_{gk1894}, and P_{gk2150} have been identified as being inducible by maltose, lactose, myo-inositol, and D-galactose, respectively (Suzuki *et al.*, 2013b). Notably, expression from promoter P_{gk704} increased 4.5- and 12-fold by addition of either soluble starch or maltose to culture medium. Growth on the pentose sugars D-xylose or L-arabinose inhibited gene expression but there were negligible effects on expression during growth on D-glucose, D-galactose, sucrose, melibiose, lactose, myo-inositol, cellobiose, or fructose. The promoter $P_{\beta glu}$, isolated from the cellobiose-specific phosphotransferase system (PTS) operon of *G. thermoglucosidasius* NCIMB11955, enhanced expression of *pheB* in the presence of cellobiose, but was also activated by glucose and xylose (Bartosiak-Jentys *et al.*, 2013). The native promoter of the sucrose-utilisation operon in *G. stearothermophilus* NUB3621, P_{surP}, has been demonstrated to increase α-galactosidase expression by 5-fold in the presence of sucrose, although characterisation in the presence of other sugars has not been reported (Blanchard *et al.*, 2014).

Alternatively, promoters can be utilised that provide extremely high levels of protein expression under aerobic growth conditions, but are switched off once the cell culture reaches late-exponential phase. Therefore, glycosyl hydrolase expression would only occur for early stages of cell growth and silenced in the latter stages of growth when the fermentation organism ferments the desired organic product. To date, no such anaerobically-silenced promoter functional in *Geobacillus* spp. has been characterised. However, due to the recent focus on engineering fermentation pathways in *Geobacillus* spp*,* promoters that can be induced under anaerobic conditions have been reported.

In particular, promoters of the lactate dehydrogenase genes, P_{ldh}, isolated from *G. stearothermophilus* NCA1503 and *G. thermodenitrificans* DSM 465^T have been characterised and applied in the production of ethanol and isobutanol, respectively (Lin *et al.*, 2014, Cripps *et al.*, 2009). However, it has been suggested that P_{ldh} activity is more active under oxygen limiting conditions than during fully anaerobic growth, and potentially induced as a result of the transitory change in redox conditions between aerobic and anaerobic growth (Bartosiak-Jentys *et al.*, 2013).

Therefore, although promoters have been reported that function in *Geobacillus* spp., there remains a continuing demand for improvements in the *Geobacillus* spp. expression toolkit, specifically the need for strong constitutive promoters, as well as strongly-inducible promoters with low basal activities.

1.10 Secretion

The utilization of polysaccharides by a CBP organism requires the secretion of expressed glycosyl hydrolases into the extracellular milieu. Understanding and control of this process is therefore necessary for an effective CBP system. Moreover, the controlled secretion of heterologous proteins could pave the way for the use of *G. thermoglucosidasius* as an expression host for commercially useful proteins (eg. lipases, amylases), since protein secretion facilitates simpler protein recovery.

In order for bacterial proteins to be secreted, they are synthesised as precursors with an N-terminal signal peptide (Heijne, 1990) that, once recognised by soluble targeting factors, facilitates the protein's transport to the secretion machinery in the cell membrane (Driessen *et al.*, 2001). Although signal peptides do share distinct features with common characteristic properties, there is negligible consensus found between them at the primary sequence level (Bendtsen *et al.*, 2004) **(Figure 6)**. These include a positively charged N-terminal domain followed by hydrophobic domain that gets retained in the cell membrane. This is then followed by a hydrophilic domain containing a signal peptidase (SPase) recognition site. As the protein is being translocated from the cytoplasm, across the hydrophobic membrane and into the extracellular milieu, the native SPase cleaves at its recognition site, releasing the mature protein, which folds extracellularly. The remaining signal peptide is subsequently degraded.

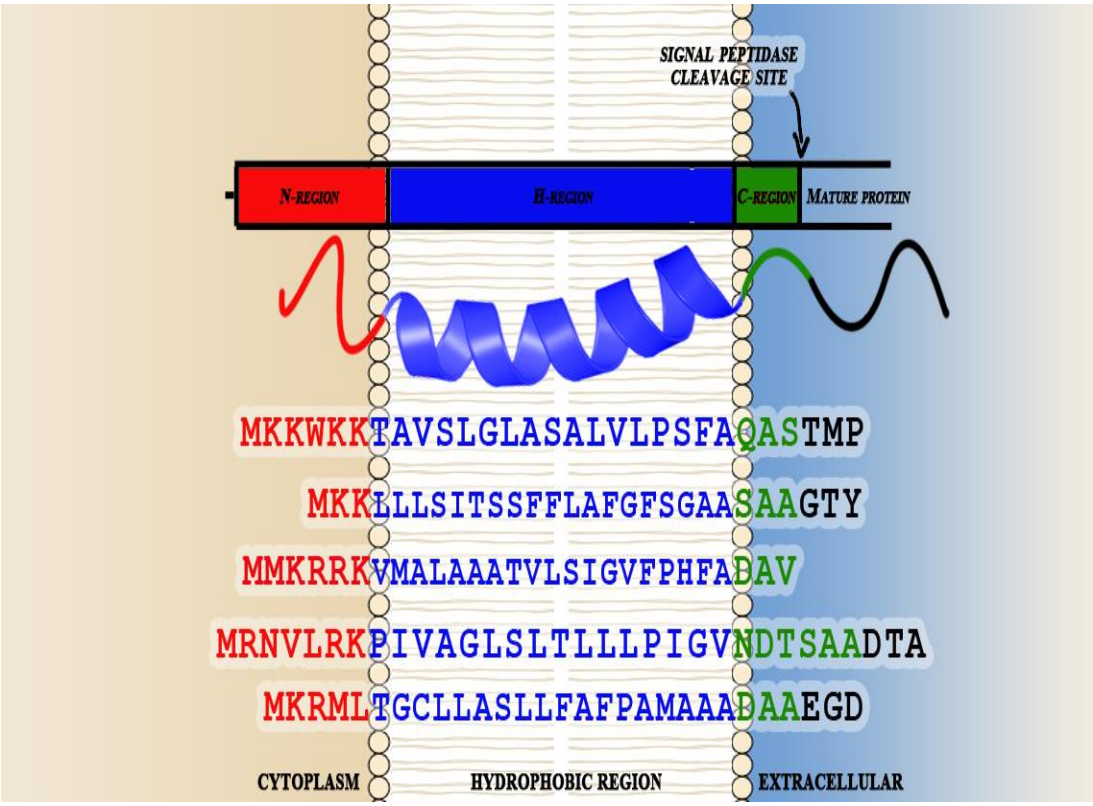


Figure 6 – Common features of N-terminal signal peptides. Signal peptides of secretion proteins contain short positively charged N termini (N region) and a membrane spanning hydrophobic region (H-region) followed by a signal peptidase cleavage site (C-region)

To date, there is limited experimental data on protein secretion in *Geobacillus thermoglucosidasius*, although annotations of the *G. thermoglucosidasius* C56-YS93, TM242 and TNO-0.90 genomes show elements of conventional bacterial secretion pathways.

The Sec Pathway

The major bacterial secretion pathway is the Sec pathway (Palmer *et al.*, 2010), which is involved in exporting proteins into the surrounding milieu in an unfolded state. In fact, the Sec pathway is found in all three domains of life, and is essential in protein transport across eukaryotic organelle membranes (Stephenson, 2005). Nevertheless, the Sec pathway is subdivided into co-translational secretion of proteins and post-translational secretion of proteins, both mediated by the recognition of N-terminal signal peptides. For decades, the general scientific consensus differentiated co-translational secretion and post-translational secretion based on recognition of the signal peptides by either the Secretion Recognition Particle (SRP) protein **(Figure 7)** or the protein SecB, respectively. However, recent studies in *E. coli* have reported that, whereas thousands of inner membrane proteins are targeted by SRP, only a handful of secretory proteins are targeted. In fact, a study has exploited the

rapidly-folding thioredoxin protein as a reporter to identify a subset of *E. coli* signal peptide sequences that promote signal recognition particle-dependent translocation.

In the current model, the vast majority of bacterial secretory proteins reach the translocon complex post-translationally, after more than two-thirds of the polypeptide chain has been synthesized at the ribosome. Proteins for secretion can be targeted to the channel either by binding to cytoplasmically diffusing SecA, or to SecY-bound SecA with or without assistance from chaperones like trigger factor (TF) or SecB, which is absent in most Gram-positive bacteria.

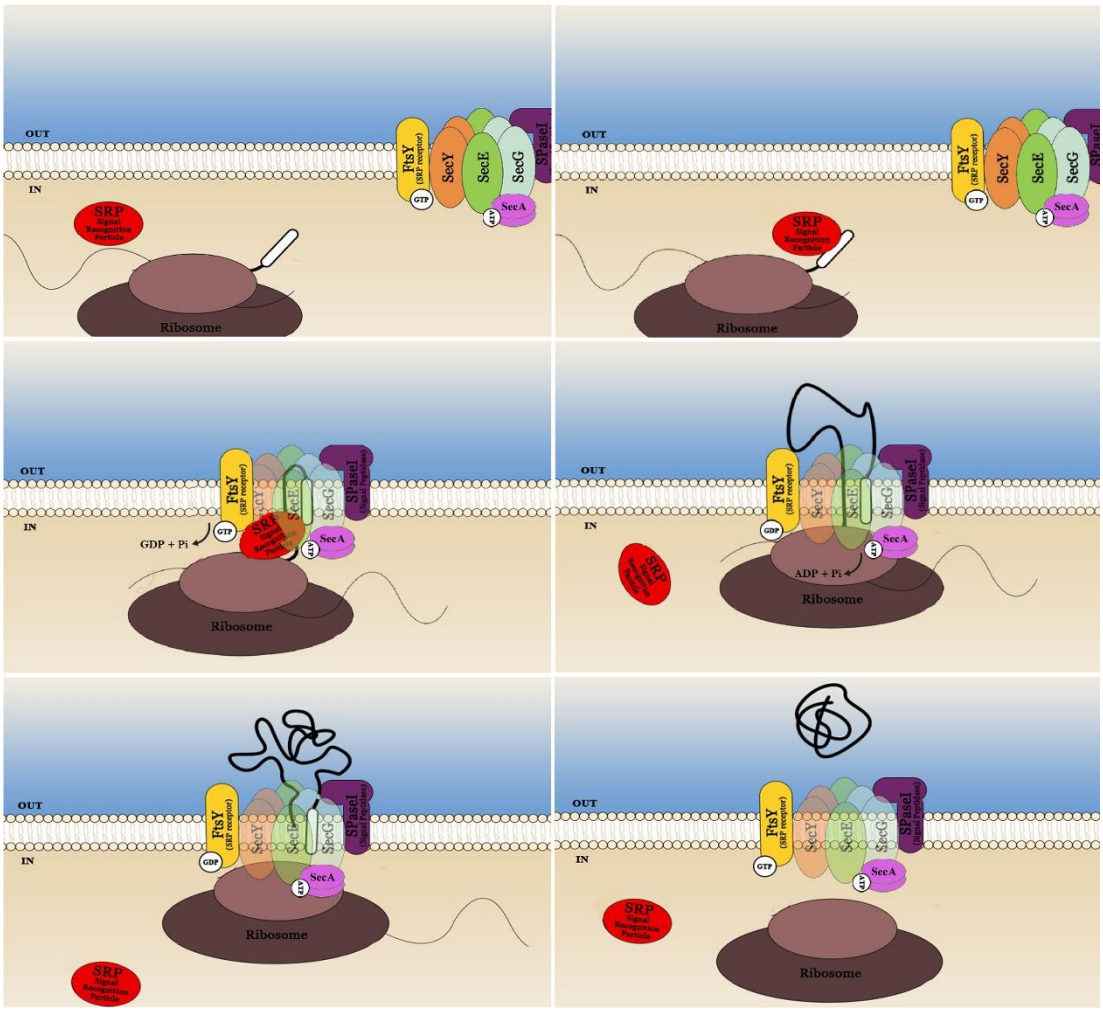


Figure 7 – The bacterial Sec-SRP Pathway. SRP recognises a signal peptide nascent polypeptide **(1-2)** and targets the polypeptide-ribosome complex to the cell membrane, where it docks with its membrane-bound receptor, FtsY **(3)**. FtsY subsequently interacts with the translocation motor SecA at the membrane translocation site (Papanikou *et al.*, 2007); a proteinacious channel in the membrane composed of proteins SecYEG. Continuous cycles of ATP binding and hydrolysis by SecA energise the system to drive the translocation of the targeted protein through the SecYEG translocon **(4)**. For

post-translational secretion, either SecA, SecB or TF recognise the signal peptide and maintain a loosely folded conformation of the synthesised protein as it is targeted to the translocon, where SecA facilitates the previously described translocation process. During translocation, the signal peptide is retained within the cell membrane **(3-5)** and the mature protein released by cleavage at the SPase recognition site **(6)** as described in Fig 6 (de Keyzer *et al.*, 2003).

Protein-protein interaction studies have revealed the presence of further intrinsic components of the Sec pathway, including SecDF, YajC and YidC (Yamane *et al.*, 2004). Although not essential for the release of mature protein from the membrane, *B. subtilis* SecDF has been shown to be required to maintain a high capacity for protein secretion (Bolhuis *et al.*, 1998). YidC has been shown to mediate membrane protein insertion, but intriguingly has been reported to be able to act independently from the Sec pathway (Serek *et al.*, 2004), interacting directly with the ribosome (Kedrov *et al.*, 2013), and forming a heterotetrameric complex with SecDF and YajC (Nouwen and Driessen, 2002)

The Twin Arginine Targeting (TAT) system

The Twin Arginine Translocation (TAT) pathway facilitates protein secretion while keeping proteins in a folded state (Kanehisa and Goto, 2000). This allows proteins to acquire and bind complex cofactors prior to transportation across the cytoplasmic membrane **(Figure 8)**. Cofactors that have been characterised as associated with TAT-dependent transport are divided into those that contain a nucleotide moiety and metal–sulphur clusters (Palmer *et al.*, 2005). Since metal ions compete for binding sites in proteins, the TAT system has been reported to transport proteins that require intracellular acquisition of metal ions under controlled conditions (Aldridge *et al.*, 2008). Intriguingly, the TAT system has been reported to facilitate the transportation of hetero-oligomeric complexes, which have formed intracellularly, across the cell membrane after recognition of a signal peptide in one of the constituent subunits (Sauve *et al.*, 2007).

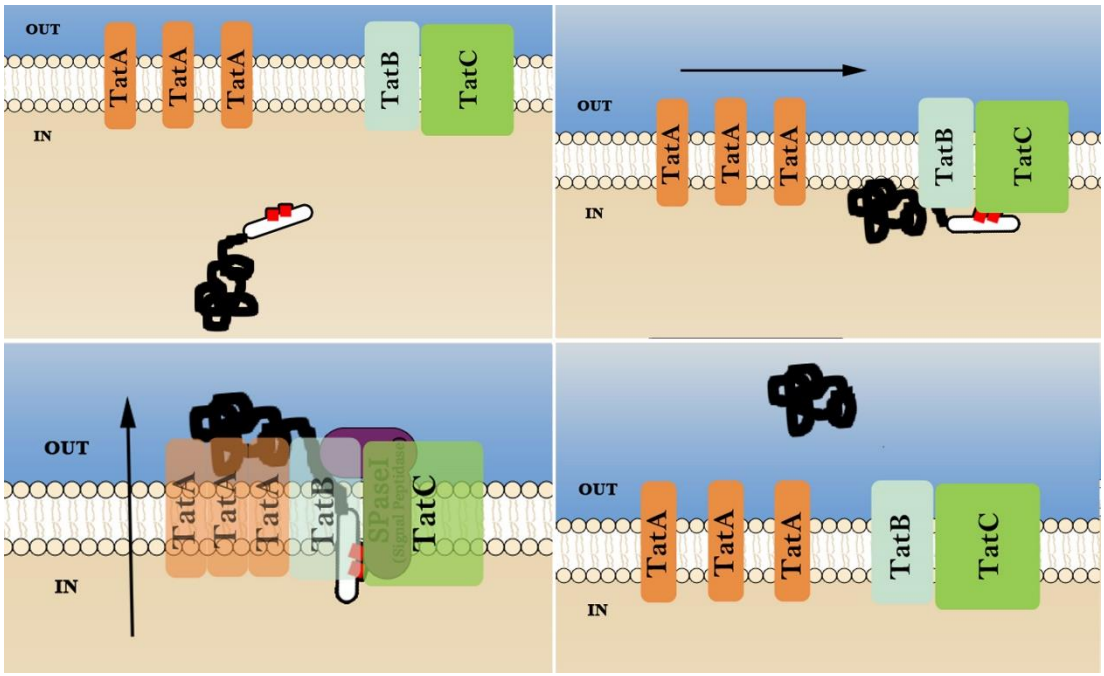


Figure 8 – The Tat Pathway of folded protein translocation. Multiple copies of TatA monomers are initially present as dispersed protomers and TatB and TatC are associated as a complex **(1)**. The TAT-targeting signal peptide is recognised and bound by the TatBC complex in an energy-independent process; the twin-arginine (RR) motif is specifically recognised by a site in TatC **(2)**. The TatA protomers are then recruited to the TatBC complex forming an active TatABC translocation complex, and the passenger domain of the secreted protein crosses the membrane via the polymerised TatA component. The signal peptide is normally proteolytically removed by the signal peptidase on the extracellular face of the cell membrane **(3)** releasing the secreted protein and dissociating the TatA protomers to free protomers **(4)**.

Current knowledge of *Geobacillus* spp. secretion

Recently, the secretion of a *G. stearothermophilus* α-amylase and a truncated-cellulase from *Pyrococcus horikoshii*, has been achieved with *G. kaustophilus* HTA426 using their native signal sequences (Suzuki *et al.*, 2013b).

However, evidence from the closely-related *Bacillus* genus suggests that no single "one-size-fits-all" signal peptide exists for the optimal secretion of different proteins. *Brockmeier et al* demonstrated that the high secretion rate of a heterologous cutinase reporter in *B. subtilis* conferred by one signal peptide was not reproduced when the secreted protein was changed to a lipase **(Figure 9)** (Brockmeier *et al.*, 2006). The hypothesis for this phenomenon centres on the concept that a signal peptide and secreted protein interact, constituting a unique unit, where the N-terminus of the mature protein and C-region of a signal peptide have been co-evolved as a 'signal peptide-mature protein' junction. As a consequence, high level secretion

of heterologous proteins can be limited by an ineffective pairing of signal peptide and mature protein. Assuming this phenomenon to be true in *G. thermoglucosidasius*, it would result in dramatically reduced secretion rates of bioprocess-relevant enzymes if only a single signal peptide was used.

Therefore, for optimisation of heterologous protein secretion in *Geobacillus*, a secretion signal peptide library must be constructed and a simple system devised for the screening of optimal signal peptide partners for a given heterologous protein.

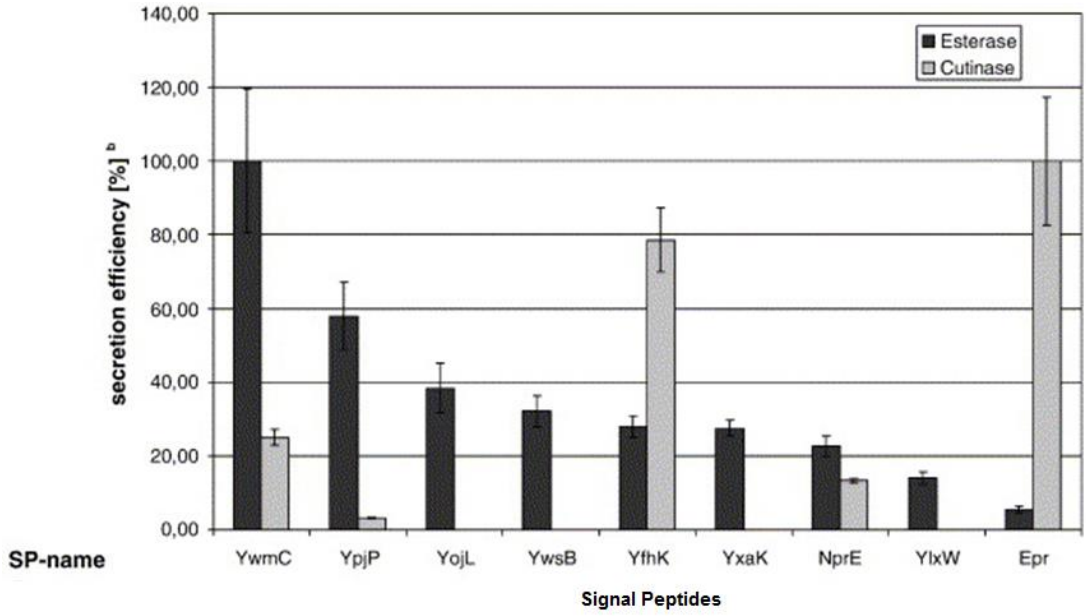


Figure 9 – Identification of the most efficient signal peptide for secretion of the heterologous esterase EstCL1 in *B. subtilis* (Adapted from (Brockmeier *et al.*, 2006)). A signal peptide library containing a mixture of all Sec-type signal peptides was fused to the metagenomic esterase EstCL1 and culture supernatants of 1000 *B. subtilis* TEB1030 transformants were screened for secreted esterase activity (black). Plasmid DNA of the ten clones that showed highest activity were sequenced and directly compared to previously-measured efficiencies of the signal peptide in cutinase secretion (grey).

1.11 The tripartite pUCG3.8 expression system

Due to the combinatorial nature of heterologous protein expression, secretion and glycosyl hydrolase composition in an effective CBP system, the Leak group created a modular system to bring these various genetic component parts together in different combinations. This has

been previously demonstrated by the heterologous expression and secretion of a *Thermotoga maritima* MSB8 glycosyl hydrolase in *G. thermoglucosidasius* NCIMB11955 using a signal peptide from a *G. thermoglucosidasius* C56-YS93 xylanase (Bartosiak-Jentys *et al.*, 2013). As shown in

Figure 10, this combinatorial expression cassette, henceforth referred to as the pUCGXXX expression system, is achieved by the universal use of four restriction sites.

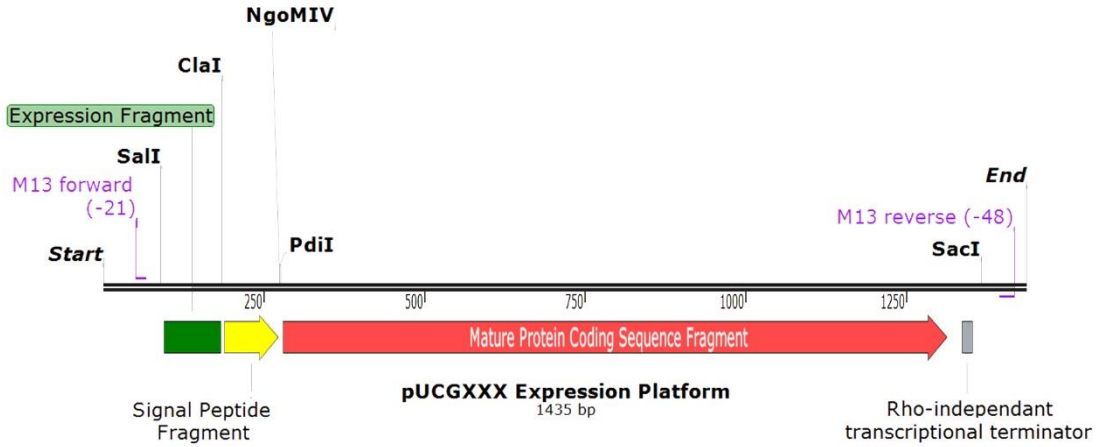


Figure 10 – The pUCGXXX expression cassette.

Promoters

Promoters function as transcription activators and binding sites for the native RNA polymerase. However, the expression of heterologous proteins is additionally influenced by the strength of the ribosomal-binding site (RBS) that is downstream of the promoter, which facilitates ribosomal recognition of the protein-encoding mRNA. Moreover, the 5' untranslated region, found between the transcriptional start site (TSS) of a promoter and the ATG translational start codon, has been shown to have a strong influence on mRNA stability and resistance to intracellular RNAses, further increasing protein expression rates. Therefore, in contrast to other expression systems, such as the BglBricks genetic module system (Anderson *et al.*, 2010), the pUCGXXX expression system involves the combined transfer of promoter-RBS regions in a single expression fragment.

In order to change the expression fragment of an assembled pUCGXXX expression construct, the promoter region of interest must be amplified with an attached *Sal*I site to the 5' end and a *Cla*I site to the 3' end. The *Sal*I site facilitates the insertion of the expression fragment at the 5' end of the pUC multiple-cloning-site, and must be upstream of all essential promoter elements, such as the -10 box, the -35 box and other regulatory binding elements; these can

be predicted by promoter prediction tools, such as BPROM (Solovyev and Salamov, 2011). The last two base-pairs (AT) of the 3' end *Cla*I site (ATCGAT) are the first two nucleotides in the ATG start codon, which ensures that expression fragments are precisely aligned to the ATG start codon of all downstream genes.

Signal Peptides

For production of extracellular proteins, such as the glycosyl hydrolases examined in this project, these proteins require an N-terminal signal peptide which is recognised by the *Geobacillus* secretion machinery. Signal peptide-coding sequences are relatively similar in size, not normally exceeding 160 bp.

In order to change the signal peptide encoding region of an assembled pUCGXXX expression construct, the region must be amplified with an ClaI site at the 5' end and a *Pdi*I/*Ngo*MIV site at the 3' end. As mentioned above, the last two base-pairs (AT) of the 5' end *Cla*I site (ATCGAT) form the first two nucleotides in the ATG start codon, which ensures that the signal peptide fragments is precisely aligned downstream of any given expression fragment. The 3' *Pdi*I/*Ngo*MIV site (GCCGGC) facilitates cleavage with either of two distinct restriction enzymes. *Pdi*I cleaves the site with GCC'GGC activity, producing blunt-ends, whereas *NgoM*IV cleaves the site with G'CCGGC activity, producing 4-nt 5' overhangs. Blunt-end cloning via the *Pdi*I-mediated digestion facilitates the sub-cloning of blunt-end signal peptide encoding and gene fragments, and was utilised in previous versions of the pUCGXXX system. However, sticky-end cloning using *Ngo*MIV-mediated digestion has largely superseded this, since the 4-nt 5' overhang (CCGG) provides strong, discriminatory hydrogen-bonding capabilities.

The *Pdi*I/*Ngo*MIV site is also compatible with signal peptide characteristics. A high percentage of signal peptide carboxyl-terminal ends contain an alanine residue at the -1 position (Von Heijne, 1998), which is encoded by the GCN codon. Therefore, the GCC codon at the beginning of the restriction site can be used to replace the final GCN codon of the amplified signal peptide. The following GGC codon of the *Pdi*I/*Ngo*MIV site translates to a small glycine residue, which is present at the amino-terminus of all heterologous secretion proteins using this system.

Genes and Open Reading Frames Mature Protein Coding Sequence Fragments (ORFs)

The amplification of mature-protein coding sequences (MCS) fragments of the secreted protein to be investigated requires the identification and removal of the native signal peptide encoding region. The MCS fragment must be amplified with an attached *Pdi*I/*Ngo*MIV site at the 5' end, a transcriptional terminator after the native stop codon, and a *Sac*I site at the 3' end. However, for intracellular proteins this simply requires a 5' *Cla*I site, a transcriptional terminator after the native gene stop codon, and a *Sac*I site at the 3' end. In both cases, the *Sac*I site facilitates the insertion of the MCS at the 3' end of the pUC multiple-cloning-site.

The universality of the pUCGXXX system facilitates the seamless swapping of expression fragments by the utilisation of *Sal*I and *Cla*I sites to modulate the expression rates of a given heterologous gene. Additionally, the *Cla*I and *Pdi*I/*Ngo*MIV sites can be exploited for the screening of multiple signal peptides for optimal secretion of a protein of interest. Lastly, the insertion of an intracellular or extracellular gene of interest for characterisation in *Geobacillus* spp. can be mediated by either *Cla*I/*Sac*I restriction enzymes or *Ngo*MIV/*Sac*I restriction enzymes respectively.

1.12 Project Overview

This central focus of this project is the development, application and evaluation of strategies to convert *G. thermoglucosidasius* TM242 (or its parent NCIMB 11955) into an organism capable of CBP, which requires a number of engineered steps **(Figure 11)**.

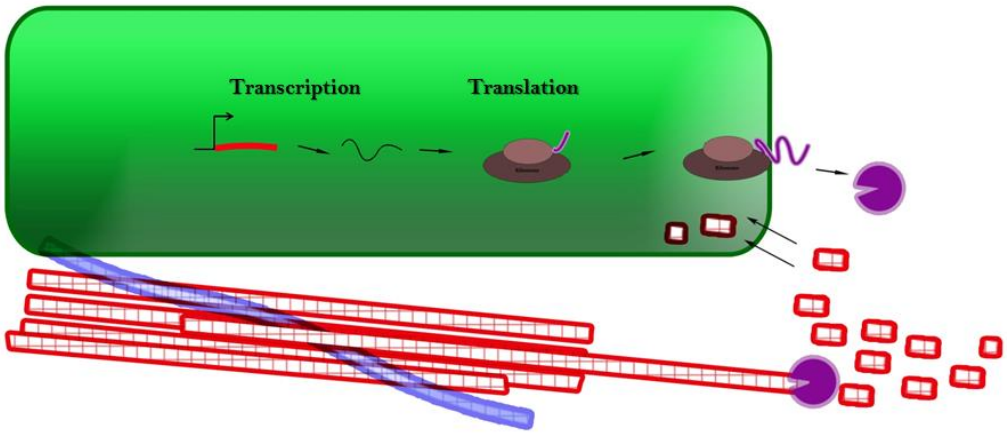


Figure 11 – Fundamentals of Strategy 2 CBP Engineering. To produce high levels of hydrolytic activity, *G. thermoglucosidasius* must be engineered for strong constitutive or inducible expression of glycosyl hydrolases at both the transcriptional level, with strong promoters (1), and at the translational level (2), with strong ribosomal binding sites. The optimisation of protein secretion (3) is important to

ensure efficient export of the expressed enzymes. Since *G. thermoglucosidasius* grows optimally at elevated temperatures (60 °C), these glycosyl hydrolases (purple sector (4)) need to actively degrade cellulose (red) and hemicellulose (blue) at elevated temperatures.

Phase I

As there was limited data available in the literature on the expression and secretion capabilities of *G. thermoglucosidasius*, Phase I of the CBP project involved the *in silico* identification and subsequent characterisation of putative promoters, signal peptides and utilisation capabilities of *G. thermoglucosidasius* and other *Geobacillus* spp. In addition, since development of *G. thermoglucosidasius* TM242 into a CBP organism would require the combinatorial expression of multiple glycosyl hydrolases, further development of the pUCGXXX system was required.

Phase II

To engineer cellulolytic capabilities in *G. thermoglucosidasius,* Phase II involved the creation of a panel of glycosyl hydrolase-secreting *G. thermoglucosidasius* strains and the subsequent analysis of the enzyme activities in their culture supernatants. Since cell surface attached cellulolytic complexes are of interest to the scientific community, strategies were also evaluated for the display of cellulases on the cell surface.

Phase III

The rate of enzyme secretion is a potential rate-limiting step in maximising the extracellular cellulolytic activity. Therefore, expressed cellulases required an optimised secretion signal for effective secretion. The demonstrated absence of a "one-size-fits-all" signal peptide for a broad range of mature proteins in *B. subtilis* **(Figure 9)** illustrated the requirement for screening a *G. thermoglucosidasius* signal peptide library to determine the optimum combination of signal peptide and secreted MCS. Phase III therefore involved the screening for optimal secretion of the glycosyl hydrolases of interest chosen from Phase II.

Phase IV

The broader objective of this project is the evaluation of the feasibility of CBP as a commercial process for the production of organic products from lignocellulosic biomass. To address this, Phase IV involved the combinatorial analysis of different co-cultures of glycosyl hydrolase-secreting *G. thermoglucosidasius* against either untreated or pre-treated *Miscanthus* x

giganteus. Since different plants have different lignocellulosic component ratios and hemicellulose branch signatures, this combinatorial analysis can be extended to other second generation feedstocks, such as sugar beet pulp, dried distiller's grains and solids (DDGS), municipal solid waste and more.

Chapter 2: Materials and Methods

2.1 Solutions, media, buffers and gels

All chemicals were obtained from Sigma-Aldrich Company Ltd (Poole, UK), Merck Chemicals Ltd (Nottingham, UK) or BDH Laboratory Supplies Ltd (Poole, UK)

Preparation	Components
DNA Agarose gel	0.8% agarose; 1x TAE buffer
TAE (50x)	2M Tris-HCl (pH 7.6); 5.7% glacial acetic acid; 500mM NaEDTA
***Geobacillus* spp. Electroporation buffer**	0.5M sorbitol; 0.5% mannitol; 10% glycerol
***Escherichia coli* TSS buffer**	10% PEG 8000; 30mM $MgCl_2$; 5% DMSO; LB medium
2TY medium	1% yeast extract; 1.6% tryptone; 0.5% NaCl
Luria-Bertani (LB) medium	0.5% yeast extract; 1% tryptone; 1% NaCl
TGP medium	1.7% tryptone; 0.3% soy peptone; 0.25% K_2HPO_4; 0.5%NaCl; 0.4% sodium pyruvate; 0.4% glycerol.
ASM ≠	8 mM citric acid, 5 mM $MgSO_4$, 20 mM NaH_2PO4, 10 mM K_2SO_4, 25 mM $(NH_4)_2SO_4$, 80 µM $CaCl_2$, 1.65 µM Na_2MoO_4, 5 ml/L Trace Element Solution*
Trace Element Solution	1.44 g/L $ZnSO_4.7H_2O$, 0.56 g/L $CoSO_4.6H_2O$, 0.25 g/L $CuSO_4.5H_2O$, 5.56 g/L $FeSO_4.6H_2O$, 0.89 g/L $NiSO_4.6H_2O$, 1.69 g/L $MnSO_4$, 0.08 g/L H_3BO_3, 60 mM H_2SO_4
Water	Purified using twin-bed deioniser (Purite, UK)

≠ These recipes do not contain a carbon Source. If a specific carbon source is required, then glucose, yeast extract, tryptone, or essential amino acids can be added..

2.2 Bacterial Strains

- ***Escherichia coli*** **JM109** (Norrander *et al.*, 1983). Genotype: *end*A1*,rec*A1*,gyr*A96*,thi,hsd*R17 *(*rk-,mk+*),rel*A1*,sup*E44*,Δ(lac-pro*AB*), [*F´*tra*D36*,pro*AB*, laq*IqZΔM15*].*

  This strain is a host for general purpose cloning and plasmid propagation. It is recombinase (*rec*A) and endonuclease (*end*A) deficient, ensuring DNA stability and gives high-quality plasmid preparations. These cells are also deficient in β-galactosidase activity due to deletions in both genomic and episomal copies of the *lac*Z gene.

- ***Escherichia coli*** **Bioblue®** (Bioline, London, UK). Genotype: *rec*A1, *end*A1, *gyr*A96, *thi*-1, *hsd*R17(r_k-, m_k+), *sup*E44, *rel*A1, *lac* [F' *pro*AB*lac*I^qZΔM15 *Tn*10(Tet^r)]

  This strain is a host for general purpose cloning and plasmid propagation. Similar to *E. coli* JM109, this strain is recombinase (*rec*A) and endonuclease (*end*A) deficient, ensuring DNA stability and gives high-quality plasmid preparations. These cells are also deficient in β-galactosidase activity due to deletions in both genomic and episomal copies of the *lac*Z gene.

- ***Geobacillus thermoglucosidasius*** **TM242** (Cripps *et al.*, 2009). Genotype: *ldh*A$^-$*pfl*$^-$P_*ldh*/*pdh*^{up}

  This strain was originally supplied by TMO Renewables Ltd, and is a high yield ethanol producing mutant of the wild-type *G. thermoglucosidasius* NCIMB 11955 strain described in **section 1.6**.

- ***G. thermoglucosidasius*** **C56-YS93.** Genotype: wild-type strain.

  This strain was originally supplied by Dr David Mead from Lucigen Corporation.

Cell culture

Escherichia coli JM109 and BioBlue strains were used as hosts for plasmid propagation during molecular biology procedures. *G. thermoglucosidasius* strains were used as expression hosts or to propagate sufficient biomass for the purification of genomic DNA.

E. coli cells were typically grown in Luria Broth medium at 37°C in either 50 ml centrifuge tubes (up to 10 ml cultures) or 250 ml baffled conical flasks (greater than 10 ml cultures), with shaking at 225 rpm. *G. thermoglucosidasius* strains were typically grown in 2TY or TGP media between 52-60°C with shaking at 225 rpm. Antibiotics were used at the following concentrations: ampicillin at 100 µg/ml, and kanamycin at 50 µg/ml for *E. coli* cells

and 12.5 µg/ml for *G. thermoglucosidasius* cells. Constant temperature, mixing and aeration in liquid culture were achieved using an *Innova44* shaking incubator (*New Brunswick Scientific*).

Strain storage

To preserve bacterial transformants harbouring expression vectors with verified target gene insert sequences, glycerol stocks were prepared and stored at -80°C in cryogenic vials. Aliquots of 1.4 mL of overnight cultures, in the required growth medium, were supplemented with sterile glycerol to a final concentration of 20% (v/v). These stocks were mixed thoroughly and snap-frozen using liquid nitrogen or dry ice prior to immediate transfer to storage.

Quantification of bacterial cell density

Duplicate 200 µl cell culture samples were added to a 96-well plate (*Costar Corning*) and the absorbance at 600nm (OD_{600}) measured in a Synergy HT plate reader *(Bio-Tek*). If the OD_{600} exceeded 0.9 absorbance units (AUs), ¼ dilutions of cell culture samples were made and measured.

Preparation and Transformation of chemically competent *E. coli*

Chemically competent *E. coli* cells were prepared from an overnight seed culture, according to the method described in (Chung *et al.*, 1989). In short, 25-50ml of fresh LB media in a 200ml conical flask was used to grow the required *E. coli* strain to an OD_{600} of 0.2 - 0.5 AU. Once split into two 50 ml centrifuge tubes, the culture was incubated on ice for 10 min, followed by centrifugation for 10 minutes at 3000 rpm (4°C). The supernatants were removed, and cell pellets were resuspended in ice cold TSS buffer to 10% of the original culture volume. 100 µl aliquots were subsequently distributeded into chilled microcentrifuge tubes and stored at -80°C.

For the transformation of plasmid constructs smaller than 9,000 bp in size, these chemically competent *E. coli* cells were transformed according to the classic 42°C heat-shock method optimised in (Inoue *et al.*, 1990).

Transformation of *E. coli* JM109 by electroporation

The transfer of large plasmid constructs, greater than 9,000 bp in size, was attempted by transformation of electrocompetent JM109 cells with column-purified ligation reaction mixtures. 0.5 µL of purified DNA was added to 40 µL of electrocompetent cells and incubated on ice for 5 min. These cells were then transferred into chilled 1 mm path length Gene Pulser® electroporation cuvettes (Bio-Rad Laboratories Ltd, Hertfordshire, UK) and placed in the Micropulser electroporator (Bio-Rad) using the Ec1 setting for *E.coli* (5 ms electrical pulse of approximately 1.8 kV). The transformed cells were resuspended in 1 mL of 2xTY medium and incubated with shaking at 220 rpm at 37°C for 1 hour.

Preparation and Transformation of electrocompetent *G. thermoglucosidasius*

Electrocompetent *G. thermoglucosidasius* cells were prepared as described in (Taylor *et al.*, 2008), with cells processed specifically at an $OD600_{nm}$ of 1.6. The transformation protocol was further adapted from (Taylor *et al.*, 2008), with the cell recovery stage in 1 mL pre-warmed 2TY medium at 52°C and 255rpm for 1hr. The cell culture was centrifuged at 4000rpm for 5 minutes and spread onto 2TY agar plates for incubation at 50°C overnight.

2.3 Molecular Biology

All primers **(Appendix I)** were supplied as a lyophilised powder by *Eurofins Genomics* and used at a final concentration of 0.5µM during PCR.

High-fidelity amplification by Polymerase chain reaction (PCR)

To amplify promoters, signal peptides, genes, and other genetic modules of interest, PCR reactions were performed using Phusion® High-Fidelity DNA Polymerase (Thermo Fisher Scientific, Cramlington, UK) following the conditions recommended by the supplier. A standard PCR reaction (50 µL) contained 1x HF Phusion buffer, 200 µM dNTPs (dATP, dCTP, dGTP, dTTP) (Bioline), 0.25 µM of each forward and reverse specific primers, an appropriate amount of DNA template (1 pg–10 ng for plasmid DNA and 50–250 ng for genomic DNA), and 0.5 U Phusion HF DNA polymerase.

Table 3 - Temperature and time used for each PCR amplification step. Temperature X depended on the annealing temperature of the specific pair of oligonucleotides used, as calculated using the Thermoscientific oligonucleotide T_m calculator. The elongation time (Y) was dependent on the size of the target gene to be amplified (30 s per 1 kb).

	Step	Temperature	Time
	Initial Denaturation	98 °C	30 s
35 cycles	Denaturation	98 °C	10 s
	Annealing	X°C	45 s
	Elongation	72°C	Y min
	Final Elongation	72°C	10 min

The reactions were carried out in thin-walled PCR tubes placed in an Eppendorf Mastercycler® gradient PCR thermocycler (Hamburg, Germany), following the programme described in **Table 3**.

Diagnostic amplification of colony transformants by colony PCR

To identify positive clones from a large number of transformants, colony PCR reactions were performed using *Taq* DNA polymerase. This was performed either using RedTaq® (*RedTaq,* Sigma Aldrich, UK) or KAPATaq (Cambridge Biosciences, UK). A standard PCR reaction mixture (20 µL) was made by adding 0.25 µM each of the forward and reverse specific oligonucleotides to 1x Taq Master Mix, which contain the Taq DNA polymerase, Taq buffer, dNTPs and loading dye.

Genomic DNA was introduced to the PCR reaction mixture by the addition of a small amount of each selected colony. Cells were lysed and DNA template released during the initial extended denaturation step.

The reactions were carried out in thin-walled PCR tubes placed into an Eppendorf Mastercycler® gradient PCR thermocycler (Hamburg, Germany), following the programme described in **Table 4**.

Table 4 - Temperature and time used for each colony PCR amplification step. Temperature X depended on the annealing temperature of the specific pair of oligonucleotides used, as calculated by the OligoCalc oligonucleotide T_m calculator (Kibbe, 2007). The elongation time (Y) was dependent on the size of the target gene to be amplified (1 min per kb).

	Step	Temperature	Time
	Initial Denaturation	95 °C	10 min
40 cycles	Denaturation	95 °C	30 s
	Annealing	X°C	45 s
	Elongation	72°C	Y min
	Final Elongation	72°C	10 min

Restriction enzyme digests

To generate cohesive ends in the pUCG4.8 expression vectors and target expression fragment, signal peptide, gene or other genetic module, different restriction enzymes were used. All the enzymes and buffers were supplied by New England Biolabs (NEB) (Hitchin, UK) or Fermentas/ThermoScientific (Waltham, MA, USA). A standard restriction enzyme digestion reaction (20 µL) contained 2 µL of the recommended buffer, 5-10 µL of purified plasmid DNA, 0.25 U of restriction enzyme(s), and Bovine Serum Albumin (BSA) at a concentration of 100 µg/mL when required.

The reaction mixture was incubated for 1-3 h at 37°C or 65°C (depending on each enzyme optimum temperature indicated by the manufacturer). If required, a subsequent 30 minute incubation at 65°C was carried out to heat inactivate the restriction enzyme.

After restriction enzyme digests, the samples were loaded onto an agarose gel (described below) and the desired fragments purified by electrophoresis (described below).

Agarose gel electrophoresis

To purify DNA fragments produced as PCR products or restriction enzyme digests, agarose gels were prepared by dissolving 0.8-4% (w/v) agarose in TAE buffer (40 mM Tris-acetate pH 8.0, 1 mM EDTA) by heating in a microwave until boiling. The solution was then allowed to cool and a 1:10,000 dilution of SYBR safe (Invitrogen) was added. This solution was poured into a gel cassette, a comb was positioned and the gel allowed to polymerise before being placed in a gel tank and covered with TAE buffer.

DNA samples were prepared in 6x DNA loading buffer (50% (v/v) glycerol, 50 mM EDTA pH 8.0, 0.05% (w/v) bromophenol blue) and the required volume (5–30 µL) was loaded onto the gel. To determine the approximate size of the DNA, 5 µL of a 1 kb DNA ladder (Fermentas) was loaded with the samples. Electrophoresis was performed at a constant 100 V and monitored by following the progress of the dye front. Electrophoresis was continued until the DNA bands were adequately separated. The resolved fragments were subsequently visualised using a G:Box UV transilluminator (Syngene).

Gel DNA purification

To purify DNA from agarose gels, the DNA bands were carefully excised under UV light and purified using the ZR® PCR DNA Clean-Up System (ZymoResearch) following the manufacturer's instructions. An elution volume of 6-16 µL pre-warmed MilliQ water was used and the DNA was stored at -20°C.

Plasmid DNA preparation

To extract and purify plasmid DNA, cells harbouring a vector ligated with the target gene (single colonies or a loop-full of frozen glycerol stocks) were inoculated into 10 mL of 2xTY medium supplemented with ampicillin at 100 or kanamycin at 30 µg/ml. Cells were grown overnight at 37°C in a shaking incubator set to 220 rpm.

After the overnight incubation, cells were harvested by centrifugation at 6,000 x g for 20 min at room temperature. Plasmids were subsequently purified using the Qiagen® spin miniprep DNA purification system, according to the manufacturer's instructions. An elution volume of 50 µL pre-warmed MilliQ water was used and the DNA was stored at -20°C.

DNA sequencing

To verify the successful cloning of a target gene into the intended region of a specified vector, plasmid DNA was sent for sequencing to Eurofins MWG Operon (Ebersberg, Germany). The samples were sequenced, in the forward and reverse directions, using the commercially-available oligonucleotide primers required for each vector. Inserts in pJET® were sequenced using pJET2.1F and pJET2.1R primers. Inserts in pUCG-based vectors (pUCG3.8, pUCG4.8 and pUCG18) were sequenced using M13(-21)F and M13(-49)R primers.

2.4 Heterologous protein expression and analysis

Conventional restriction-ligation DNA assembly

For efficient ligation of restriction enzyme digest fragments into cloning vectors, a 3:1 insert-to-vector molar ratio was used for linear DNA fragments with compatible sticky-ends. To determine the amount of plasmid and insert to use, the DNA concentration was estimated using a NanoVue Plus spectrophotometer (GE Healthcare, Chalfont St Giles, UK).The DNA fragment mixture was subsequently added to a 10 µl ligation reaction with 1 U T4 DNA ligase (*Fermentas*) in 1% T4 DNA ligase buffer, and incubated at room temperature for 1 hour or overnight at 16°C.

Gibson Assembly

Gibson fragments were generated by PCR high-fidelity amplification using primers designed to have a 40 bp overlap between adjacent fragments. PCR products were resolved on a 1% Agarose TAE gel, visualised and purified into an elution volume of 6 µl.

5x isothermal reaction buffer:	1.33x Isothermal assembly mix:
25% PEG-8000	Taq ligase (40u/ul): 50 ul
500 mM Tris-HCl pH 7.5	5x isothermal buffer: 100 ul
50 mM MgCl2,	T5 exonuclease (1u/ul): 2 ul
50mM DTT	Phusion polymerase (2u/ul): 6.25 ul
5mM NAD	Nuclease-free water: 216.75 ul
1mM each of the four dNTPs	

Pre-prepared 15 µl aliquots of 1.33% isothermal assembly mixture, stored at -20 °C, were slowly thawed on ice until ready for use. The aliquots were stored in thin-walled PCR tubes that allow for incubation in an Eppendorf Mastercycler® gradient PCR thermocycler (Hamburg, Germany).

Purified DNA fragments were subsequently mixed in equimolar amounts at a total concentration of 20 ng/µl (total volume of >6 µl). 5 µl of this DNA mixture was then added to

the thawed 15 µl 1.33% isothermal assembly mixture, and immediately incubated at 50°C in a thermocycler for 60 min.

GoldenGate Assembly

FastDigest® *Bsa*I enzyme was used in all GoldenGate reactions (Thermoscientific, Waltham, MA, USA). A standard GoldenGate assembly reaction (20 µl) contained 2 µL of the provided FastDigest buffer, 0.5 mM ATP, 50 ng of purified GoldenGate-ready plasmid DNA, 1 U of FastDigest® *Bsa*I, and 2 U of T4 DNA Ligase (Thermoscientific, Waltham, MA, USA).

The reaction mixture was either incubated for 1-3 h at 37°C or at 16°C overnight. A subsequent 30 minute incubation at 65°C was always carried out to heat inactivate the reaction prior to transformation of competent *E. coli*.

2.5 Heterologous protein expression and analysis

Heterologous protein expression in *G. thermoglucosidasius* TM242

To express the heterologous proteins outlined in this project, a loopful of frozen glycerol stock of *G. thermoglucosidasius* TM242 harbouring the pUCG4.8-target gene plasmid, was used to inoculate 5-10 mL of pre-warmed (60°C) 2xTY medium supplemented with kanamycin. Cells were grown overnight in 50 ml centrifuge tubes at 60°C shaking at 220 rpm. Then 1 mL of this starter culture was used for inoculation in 5-10 mL (in a 50 ml centrifuge tube) or 30-50 ml (in a 500 mL baffled conical flask) of pre-warmed (60°C) 2xTY rich medium or ASM minimal medium. Growth temperature was controlled at 55-60°C with shaking at 220 rpm.

To collect supernatants, the cells were harvested by centrifugation at 6,000g for 30-60 min at 4°C. The duration of centrifugation was extended to reduce the likelihood of residual cells remaining in the assayed supernatants. The speed of centrifugation was maintained at 6,000g to reduce the likelihood of cell lysis caused by centrifugal forces. Supernatants were either filtered with a 0.22µm syringe filter unit with a hydrophilic polyethersulfone (PES) membrane that was marketed to exhibit extremely low protein binding (*Millipore*) and stored at 4°C. Alternatively, the supernatants were immediately used in enzyme assays of secreted extracellular heterologous protein (described below).

2.4 Enzyme Assays

Conventional 3,5-Dinitrosalicylic acid (DNS) Assays

Cellulase activity was assayed by measuring the amount of reducing sugars as glucose equivalents liberated by cellulose hydrolysis using DNS as described by (Miller, 1959). The substrate used was either 1.25% (w/v) carboxymethylcellulose (CMC) (Sigma-Aldrich), 1 in 1M KH_2PO_4 buffer pH 7.0. The reaction mixture, containing 1 mL of the substrate and 1 mL of enzyme-containing culture supernatant, was incubated at 60°C in a heating block for different lengths of time (15-120 minutes). The reaction was stopped by the transfer of 200 µl of reaction mixture to 400 µl DNS reagent (1% (w/v) DNS, 1% (w/v) NaOH and 0.05% (w/v) Na_2SO_3) before heating the mixture for 30 min at 100°C. After boiling was completed, the DNS processing reactions were allowed to cool for 5 min on ice. 150 µL of the DNS reactions was transferred to a 96 well microplate (Costar Corning) and absorbance was measured at 540 nm in a Synergy HT plate reader (Biotek).

A 10 mM glucose stock solution was used to make standards of known concentration (0-10mM). This standard curve was used to calculate the amount of reducing sugar as glucose equivalents liberated from the samples. One unit (U) of cellulase activity was defined as the amount of enzyme that catalyses the release of 1 µmol of reducing sugar per minute, based on the glucose calibration curve, under the specified assay conditions.

Measurements of sfGFP expression in *G. thermoglucosidasius* TM242

To characterise the sfGFP expression profiles conferred by different promoter expression constructs, a colony of *G. thermoglucosidasius* TM242 harbouring the pUCG4.8-sfGFP expression plasmid was used to inoculate 5-10 mL of pre-warmed (60°C) 2xTY medium supplemented with kanamycin. Cells were grown overnight at 60°C in a shaking incubator in a 50 ml centrifuge tube at 220 rpm. Then 100 µl of this starter culture was used as inoculum for inoculation in 5 mL (in a 50 ml centrifuge tube held at a 45° tilt) of pre-warmed (60°C) 2xTY rich medium or ASM minimal medium. Growth temperature was controlled at 55°C with shaking at 220 rpm.

To measure culture OD_{600nm}, and fluorescence of the cell cultures, duplicate 200 µl cell culture samples were transferred to a 96-well plate (*Costar Corning*) and subjected to dual analysis in a Synergy HT plate reader *(Bio-Tek*). The set program facilitates the technical-duplicate measurement of cell density at 600 nm, immediately followed by excitation of the cell samples at 480nm and measurement of emission at 510 nm. Once a culture optical

density of over 0.9 absorbance units (AUs) was reached, four fold dilutions of cell culture samples were made, measured and the reading corrected for the dilution factor.

Prospecting for glycosyl hydrolases with high activity on amorphous and crystalline cellulose

Literature databases and search engines (*Pubmed, Web of Knowledge, Google Scholar*) were searched using key phrases "cellulase/ endoglucanase/ exoglucanase/ cellobiohydrolase thermophil*" and "cellulase/ endoglucanase/ exoglucanase/ cellobiohydrolase characteris*", where * activates the wild-card feature available from these web services. Reported values for specific activities against carboxymethylcellulose, avicel and β-Barley Glucan were transferred to a Microsoft Excel spreadsheet, together with other metadata, including organism strain name, optimum growth temperature, publication reference, and conditions of characterisation reaction. Although specific activities measured in U/µmol (µmol of reducing sugar per minute per micromole of protein) are a useful measure of product turnover. However, since the majority of reports have been characterised in the conventional International Miller unit of U/mg (µmol of reducing sugar per minute per milligram of protein), values reported in U/µmol were converted using **Equation 1**. Although *Vmax* and specific activities have the same standard unit of U/mg, it must be stressed that *Vmax* is the calculated rate of reaction under substrate saturation, and caution was taken to differentiate between the two.

Equation 1:

$$\boldsymbol{Specific\ Activity}\left(\frac{\boldsymbol{U}}{\boldsymbol{mg}}\right) = \frac{\boldsymbol{Specific\ Activity}\left(\frac{\boldsymbol{U}}{\mu\boldsymbol{mol}}\right)\boldsymbol{x\ 1000}}{\boldsymbol{MW\ of\ Cellulase(Da)}}$$

Preparation of PASC

Phosphoric-acid swollen cellulose (PASC) was prepared from Avicel (Fluka, Seelze, Germany). Avicel (10 g) was dissolved in 250 ml of 99% phosphoric acid and stirred for 1 hour at 4 °C, then the suspension was diluted with 4750 ml of cold water. After stirring for 1 hour at 4 °C, the resulting amorphous cellulose was collected by filtration with 0.22 µm filter paper (Millipore, Bellerica, MA, USA). It was washed 4 times with ultrapure water, 2 times with 1% $NaHCO_3$ to neutralise the pH, and then 3 more times with ultrapure water. The cellulose paste was homogenised (2 min × 3) with a multi-speed blender (Morphy Richards,

Mexborough, UK). The resultant slurry was resuspended in 1 M KH_2PO_4 buffer (pH 7.0) for use in enzyme assays and cultures, and stored at 4 °C.

Bradford Assay

In the presence of insoluble substrates such as PASC and Avicel, , cell concentrations were measured using the Bio-Rad Protein Microassay (Bio-Rad, Hercules, CA, USA), based on the Bradford Assay. For correlation, the culture OD_{600nm} of an overnight culture of *G. thermoglucosidasius* TM242 cells, grown in 2TY rich media, was measured prior to harvesting by centrifugation at 6,000g for 10 min at 4°C. The resulting cell pellet was resuspended in 2X ASM mineral salts medium, with no added carbon source, to produce cell dilutions with measured OD_{600nm} of 1.6, 1.0, 0.2, 0.16, 0.1 and 0.02 AU. 2% w/v Avicel was subsequently added to each cell standard to reduce their culture OD_{600nm} equivalents to 0.8, 0.5, 0.1, 0.08, 0.05 and 0.01 AU. 160 μl of each cell standard was added to 40 μl of Bio-Rad Protein Assay Dye Reagent Concentrate in a 96-well plate (Costar Corning), followed by a 10 minute incubation at room temperature. Protein concentration was subsequently measured by spectrophotometry in a Synergy HT plate reader *(Bio-Tek*) set to measure at 600 nm. Background absorbance at 600nm was measured with 1X ASM mineral salts medium, and accordingly subtracted from sample measurements.

Chapter 3: The characterisation of *Geobacillus* spp. compatible promoters

Since the primary goal of this project was the heterologous expression of glycosyl hydrolases and auxiliary proteins in *G. thermoglucosidasius*, there is a fundamental requirement for strong and controlled expression platforms. As discussed in **section 1.9**, until recently, there were few characterised promoters functional in *Geobacillus* spp, and therefore developments of the *Geobacillus* protein expression toolbox are in high demand.

3.1 Construction of GoldenGate-ready superfolder GFP expression platform for characterisation of promoters in *Geobacillus* spp.

As introduced in **section 1.8**, the characterisation of promoter activity and strength under selected culture conditions involves the introduction of promoters upstream of a quantifiable expression reporter, in this case the thermostable superfolder green fluorescent protein (sfGFP).

To date, this has been achieved in the pUCGXXX system (fully introduced in **section 1.11**) by conventional restriction digest and ligation of the amplified expression fragment into a pUCG4.8_Prom_*sfGFP* vector, whereby Prom denotes any present promoter expression fragment, and *sfGFP* is the superfolder GFP gene.

However, there are a number of issues that complicate the comparison of sfGFP expression using conventionally-cloned expression constructs. Firstly, the process is time-consuming, with separate digestion of the amplified expression fragment and pUCG4.8 construct, followed by separate agarose-gel purifications, and a ligation reaction skewed in efficiency depending on insert size and sequence. Secondly, both vector and amplified expression fragment DNA is partially lost through the digestion and purification procedures. Thirdly, the use of ultra-violet light to visualise the agarose gel-resolved digestion products may introduce base substitutions. These mutations will be identified by sequencing if present in the expression fragment and sfGFP gene, but not if present in the vector backbone. If such base substitutions are present in functional regions of the vector backbone, such as the *Geobacillus* origin of replication, for example, then a potential change in plasmid copy number may result in a change in expression levels of the given promoter expression construct, skewing fluorescence assays.

To circumvent this, the GoldenGate method of cloning was exploited to construct a variant of pUCG4.8 that facilitates seamless and targeted insertion of promoter-RBS expression fragments upstream of the *sfGFP* gene[1]. The process relies on exploiting Type IIs restriction enzymes, such as *Bsa*I, which recognises the sequence 'GGTCTC', but cleaves the 6 bp sequence immediately downstream, to leave a 4-nt 5' overhang.

*Bsa*I

5'...GGTCTCN^. . . . . 3'
3'...CCAGAGNNNNN^...5'

To target an expression fragment upstream of the *sfGFP* gene in the pUCGXXX system **(Section 1.11)**, the reverse GoldenGate end consists of a *Bsa*I recognition site followed by a *Cla*I site (ATCGAT), which facilitates its insertion at the *Cla*I site of the pUCGXXX system, which partially contains the ATG start codon (ATCG**ATG**). However, the upstream *Sal*I site (GTCGAT) cannot be utilised as the universal forward GoldenGate end, since *Bsa*I digestion leaves the same 4-nt 5' TCGA overhang as that of *Cla*I site (ATCGAT). Therefore, for the targeted insertion of the expression fragment at the 5' of the pUCG4.8 multiple cloning site, the forward GoldenGate end consists of a *Bsa*I site followed by a GGTCGAC sequence. This facilitates initial GoldenGate-based insertion of the promoters, and subsequent module subcloning procedures using the flanking *Sal*I and *Cla*I sites **(Figure 12)**.

Initially, the *sf*GFP gene was amplified from pUCG18_RPLSsfGFP (Reeve, B., Imperial College London, UK) using the primer pair 1(F)/2(R), and inserted by restriction-ligation process into the *Cla*I and *Sac*I sites of pUCG4.8. To fully introduce the GoldenGate process to the pUCGXXX system, and circumvent any of the aforementioned potential mutational variance between the several expression constructs, a GoldenGate-ready section was designed and inserted between the *Sal*I and *Cla*I sites. Once digested with *Bsa*I, this construct generates the precise compatible ends for unidirectional ligation of the expression fragments containing the corresponding aforementioned GoldenGate ends. However, to facilitate this, a silent site-directed mutation of the *Bsa*I site present in the ampicillin resistance marker of pUCG4.8, was required. Using the primer pairs 1(F)/3(R) and 4(F)/5(R), the vector pUCG4.8_sfGFP was amplified in two parts (1,781-bp and 3,179-bp) and assembled by a GoldenGate reaction into pUCG4.8_GoldenGate_sfGFP.

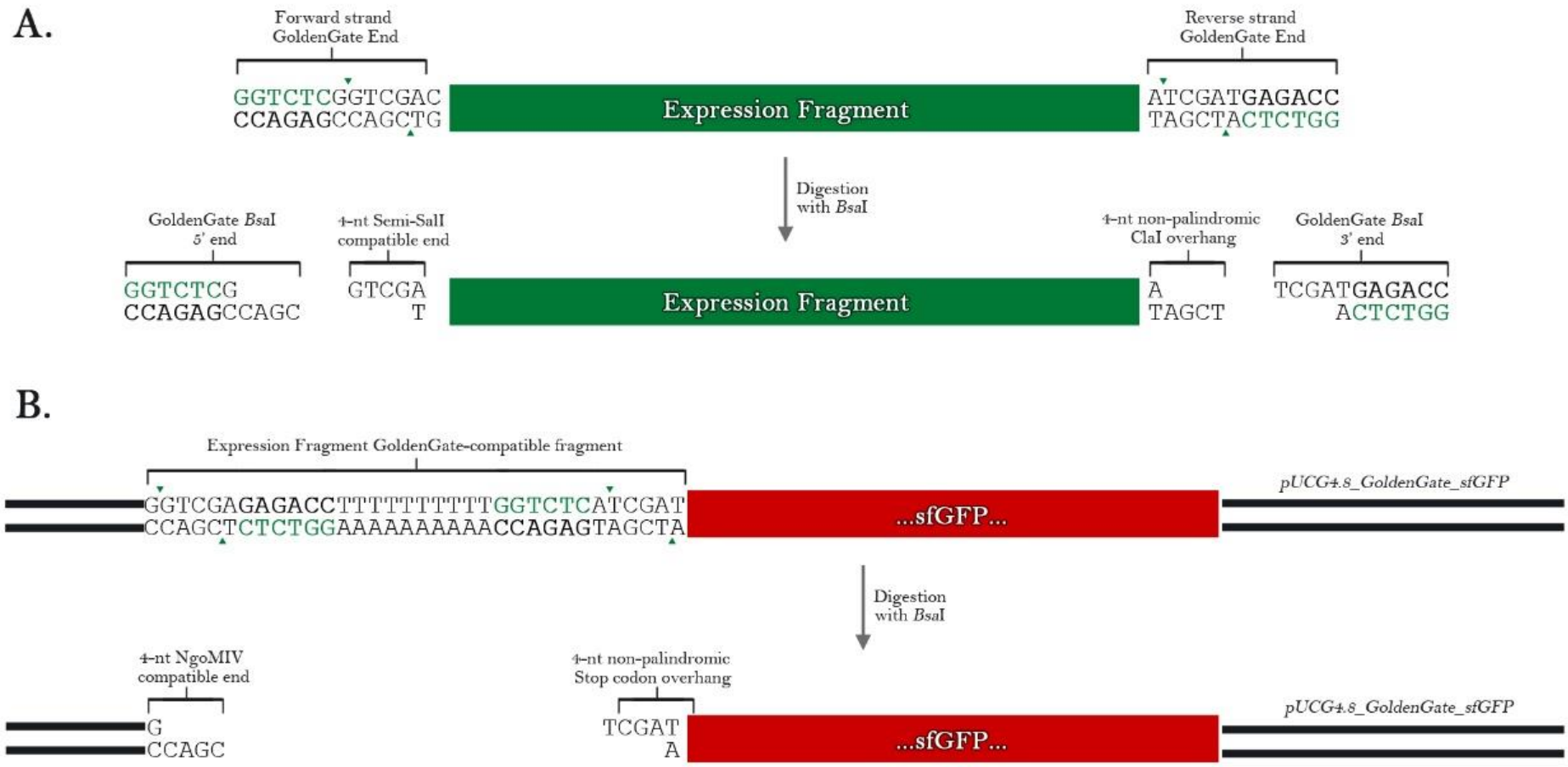


Figure 12 - GoldenGate-based insertion of expression fragments fragments into the shuttle expression vector pUCG4.8_ GoldenGate_*sfGFP*. (A) The forward and reverse GoldenGate ends that flank the amplified expression fragments are designed such that *Bsa*I digestion at the 5' *Bsa*I site (green lettering) results in the generation of specific 4-nt overhangs. (B) *Bsa*I digestion of the Expression GoldenGate-compatible fragment, which has been cloned upstream of the superfolder green fluorescent protein (sfGFP) gene, facilitates the generation of complimentary 4-nt overhangs to those generated on the amplified expression fragments. Ligation of these overhangs is simultaneously mediated by T4 DNA ligase.

It is important to note that although *Bsa*I is the commonly used Type IIs restriction enzyme, the process can be amended to utilise any Type IIs restriction enzyme that generates 5' 4-nt overhangs, such as *BsmB*I and *Bpi*I. This is simply achieved by swapping the Type IIS restriction sites and maintaining the same aforementioned forward and reverse GoldenGate ends.

3.2 Characterisation of available constitutive promoters in *G. thermoglucosidasius*

Initially, two synthetic promoters were obtained *(Ben Reeve, Imperial College London, UK)* with previously characterised constitutive activity in *G. thermoglucosidasius* TM242.

The first of these promoters was the wild-type version of the *G. thermoglucosidasius* NCIMB 11955 RplS promoter (pRPLS), which controls the expression of the most highly expressed ribosomal protein - 50S ribosomal protein L19. To further improve expression levels, the native ribosomal binding site (RBS) was replaced with the RBS of a *G. stearothermophilus* LDH gene *(Ben Reeve, Imperial College London, UK)*. As **Figure 13** shows, the replacement of the RBS results in a 2.5-fold increase in sfGFP expression, and the RplS promoter exhibits exceptionally high expression *(Ben Reeve, Imperial College London, UK)*

The second of the acquired promoters is a mutant of the uracil phosphoribosyl-transferase (UPRT) promoter (pUP2n38) demonstrated to be constitutive, but with weak-to-moderate expression. Although protein expression from pUP2n38 is likely to be insufficient for effective levels of heterologous glycosyl hydrolase expression, the promoter may be useful for the expression of proteins with auxiliary and synergistic activities to glycosyl hydrolases **(Figure 13)**.

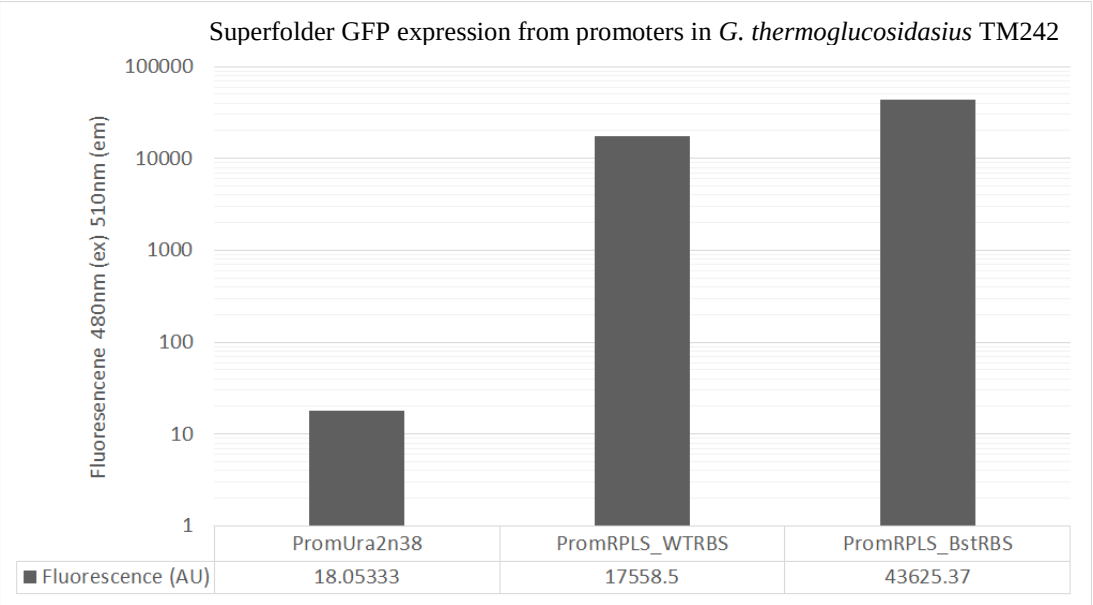


Figure 13 – Characterised activities of *Geobacillus* spp. promoters. Superfolder GFP expression levels (determined as fluorescence) conferred by a mutant UPRT promoter (PromUra2n38) and RPLS promoters with wild-type RBS (PromRPLS_WTRBS) or *G. stearothermophilus ldh* RBS (PromRPLS_BstRBS). Data from Ben Reeve, Imperial College London.

However, these promoters have been previously characterised in pUCG18-based constructs that utilise an *Xba*I site for promoter fragment insertion at the Promoter-RBS junction, whereas the present pUCG4.8 expression system utilises a *Cla*I site at the RBS-Start Codon junction. Any influences that the synthetically added *Cla*I site (ATCGAT) conveys on expression from the pRPLS or pUP2n38 promoter must therefore be investigated.

The strong and constitutive pRPLS promoter.

To investigate the strength of protein expression conferred by the pRPLS promoter in the pUCG4.8 expression platform, the promoter region was amplified from pUCG18_RPLSsfGFP using primers 6(F)/7(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method.

The expression of the construct under aerobic growth conditions was investigated in 2xTY rich media. As expected, the RplS promoter exhibited high activity under aerobic growth conditions, with superfolder GFP expression levels, calculated as GFP fluorescence divided by the present culture OD600, reaching 683 AUs/OD at a measured culture OD_{600nm} of 1.22 **(Figure 14)**.

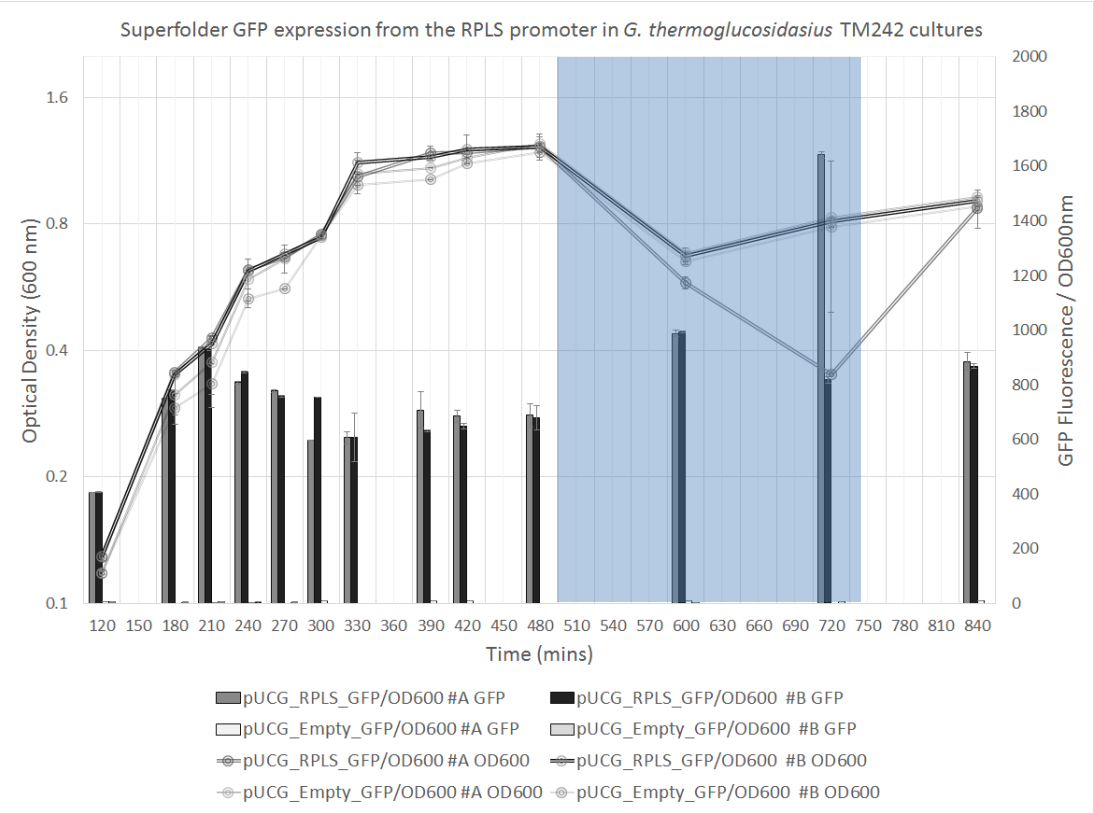


Figure 14 – Superfolder GFP expression profile from pRPLS promoter. Superfolder GFP expression levels conferred by the pRPLS_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions (clear) in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Both biological duplicates and technical duplicates were measured. Once culture OD_{600nm} remained stable, an indicator of stationary phase, a transition towards oxygen-limiting conditions was achieved (shaded) by adding further media to reduce the closed culture air volume from 90% to 30%.

Moreover, increases in sfGFP expression were observed to be correlated to culture growth rate. For example, between 300 and 330 minutes, an increase in culture OD_{600nm} by 43.9% resulted in an increase in detected sfGFP fluorescence by 30.7%, whereas between 390-420 minutes, an increase in culture OD_{600nm} by 2.2% resulted in an increase in detectable with periods of high growth rates resulting in large increases in sfGFP expression by 2.1%.

Although the further addition of culture medium to achieve oxygen limiting conditions did initially dilute culture OD_{600nm} and sfGFP fluorescence by 83% (OD_{600nm} = 0.20, GFP = 139 (data not shown)), sfGFP expression levels rose and stabilised to 619 ± 47 AUs at a measured culture OD_{600nm} of 0.625 ± 0.04 AUs, indicating that pRPLS promoter remains active under oxygen-limiting conditions.

The strong expression conferred by the pRPLS promoter, and its constitutive nature, make it the obvious candidate promoter for the heterologous expression of glycosyl hydrolases from *G. thermoglucosidasius* TM242.

The weak and constitutive pUP2n38 promoter.

To investigate the strength of protein expression conferred by the pUP2n38 promoter in the pUCG4.8 expression platform, the promoter region was amplified from pUCG18_UP2n38sfGFP using primers 8(F)/9(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method.

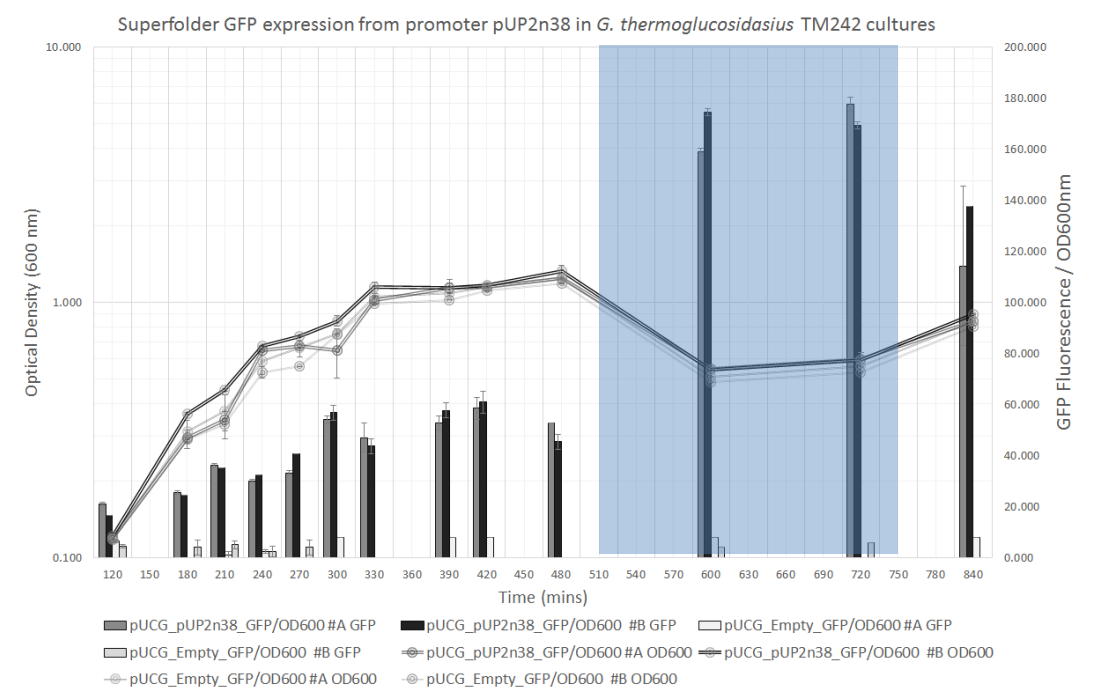


Figure 15 – Superfolder GFP expression profile from pUP2n38 promoter. Superfolder GFP expression levels conferred by the pUP2n38_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions (clear) in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Both biological duplicates and technical duplicates were measured. Once culture OD_{600nm} remained stable, an indicator of stationary phase, a transition towards oxygen-limiting conditions was achieved (shaded) by adding further media to reduce the closed culture air volume from 90% to 30%.

The expression of the construct under aerobic growth conditions was investigated in 2xTY rich media. As expected, expression of sfGFP was considerably weaker in comparison to pRPLS. For example, sfGFP fluorescence reached 69 ± 2 AUs for a culture OD_{600nm} of 1.16 AUs, which is equivalent to 8.2% of the expression conferred by the pRPLS_sfGFP constructs at the same culture OD_{600nm} **(Figure 15)**.

Nevertheless, the weak activity conferred by the pUP2n38 promoter is insufficient for effective glycosyl hydrolase expression, but remains a useful tool for the expression of synergistic proteins with ancillary activities, such as serpins and expansins **(covered in section 4.3).**

3.3 Prospection of putative promoters from transcriptomic analysis

The fundamental function of a promoter is the facilitation of RNA polymerase binding, and initiation of its corresponding gene's transcription into messenger RNA (mRNA). The characterisation of transcriptional activity can be achieved through the sequencing of all reverse-transcribed mRNA molecules from a cell culture. This whole-transcriptome analysis by next-generation sequence technologies, which is widely referred to as RNAseq analysis (Reviewed in (Wang *et al.*, 2009)).

Total RNA was isolated from duplicate aerobic and anaerobic 1.5 L batch cultures of *G. thermoglucosidasius* NCIMB 11955, the wild-type strain of the ethanologenic TM242 mutant, grown in minimal ASM media supplemented with glucose.

Table 5 presents the relative RPKM expression data of two promoters of interest that will be discussed in detail.

Table 5 – Relative RPKM expression data of two promoters of interest. Reported as the proportion of sequenced ORF transcripts (RPKM value) relative to total sequenced transcriptome (total RPKM).

		Proportion of total RPKM			
		Aerobic 1	**Aerobic 2**	**Anaerobic 1**	**Anaerobic 2**
peg3856	Acetaldehyde Dehydrogenase	0.00%	0.00%	1.35%	0.39%
peg4070	Hypothetical Protein	1.22%	0.85%	1.87%	2.47%

By utilising the powerful genome analytics software CLC Genomics *(CLC Bio)*, the forward and reverse strand FASTQ files, which are conventional FASTA sequence files with sequencing quality scores, were analysed. The two FASTQ files of each experiment were paired and aligned to all predicted ORFs from the *G. thermoglucosidasius* TM242 genome. The resultant expression values, unique gene reads, total gene reads, and RPKM values (Reads per Kilobase of transcript per Million mapped reads) of each of the 4470 annotated *G. thermoglucosidasius* TM242 predicted ORFs were exported for analysis.

Due to the variability in OD_{600nm} of the cell culture experiments, even across the replicas, it was not feasible to compare the RPKM values between aerobic and anaerobic conditions directly. Therefore, relative RPKM values were calculated, whereby the RPKM value is given as a proportion of the total RPKM value of the cell. Therefore, an increase or decrease in the expression of a gene, relative to the expression of the total transcriptome, can be considered as an upregulation of downregulation of the corresponding gene, respectively.

3.4 The exponentially-active p4070 promoter.

One ORF of particular interest, peg4070, encodes a hypothetical protein with limited sequence identity to any presently sequenced ORF. However, the RNAseq analysis has reveals considerably strong expression levels for the ORF, with sequenced peg4070 transcripts accounting for 0.9-2.5% of total transcriptomic DNA **(Table 6)**.

Table 6 – RNAseq data of *peg4070* ORF. Relative expression levels reported as the proportion of sequenced ORF transcripts (RPKM value) relative to total sequenced transcriptome (total RPKM). The average % change in RPKM values from aerobic duplicates to anaerobic duplicates was calculated based on relative RPKM values, and is presented as a product of promoter strength (average relative RPKM under aerobic conditions).

		Relative Expression Levels			
		Aerobic 1	Aerobic 2	Anaerobic 1	Anaerobic 2
		1.22%	0.85%	1.87%	2.47%
		Average % change (Aerobic → Anaerobic)		**Average % change relative to promoter strength**	
peg4070	Hypothetical Protein	116.58%		1.21%	

Although no currently characterised domains were detected in the primary amino acid sequence, the translated ORF was predicted to be a soluble secretion protein, based on prediction analysis by five prediction servers (SignalP, PrediSi, TMHMM, SosuiSignal, and LipoP). Analysis of the region upstream of the *peg4070* gene reveals the presence of strongly predicted -10 and -35 sites and putative promoter binding sites, including the OmpR binding site **(Figure 16)**. OmpR is a transcriptional repressor reported to regulate the expression of outer membrane protein F, a non-specific transport channel that allows for the passive diffusion of small, polar molecules (Forst *et al.*, 1995).

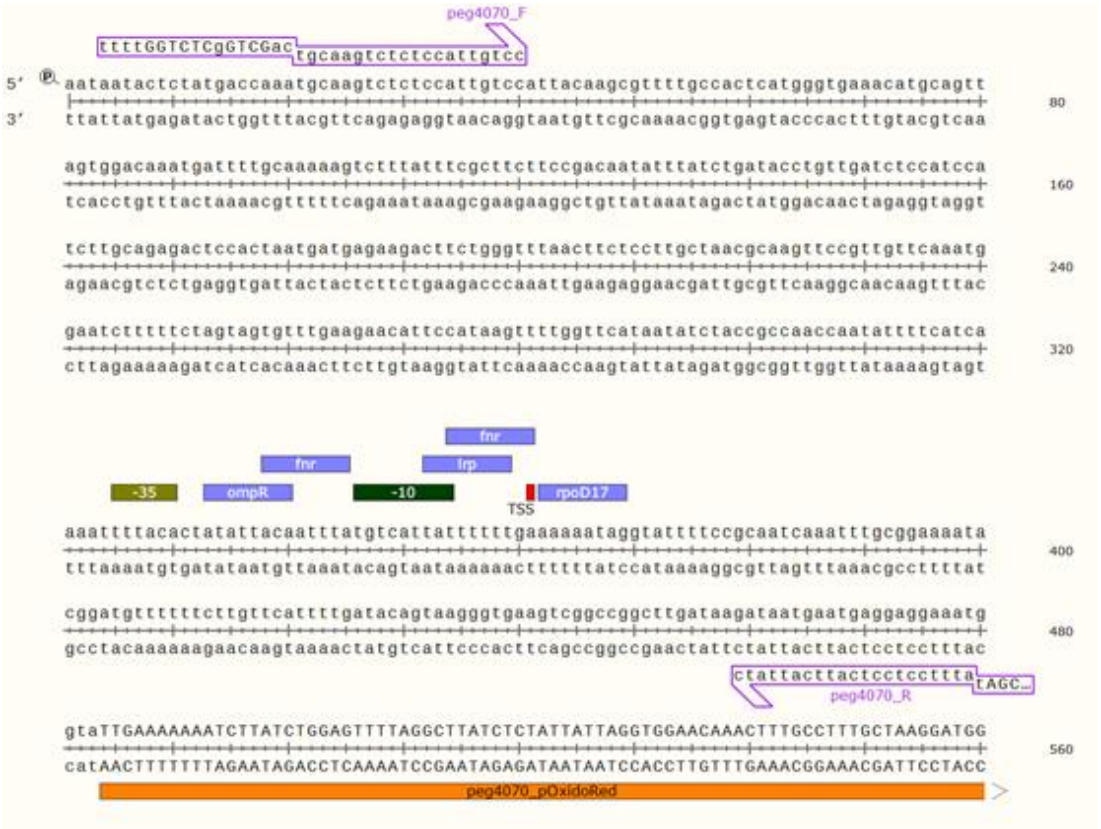


Figure 16 – Upstream promoter region of *G. thermoglucosidasius* TM242 *peg4070* gene. Promoter features were predicted using the BPROM promoter prediction server, and manually annotated using SnapGene Viewer.

Nevertheless, to investigate its transcriptional properties, the promoter region was amplified from *G. thermoglucosidasius* TM242 genomic DNA using primers 10(F)/11(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method.

The expression of the construct under aerobic growth conditions was investigated in 2xTY rich media. Consistently throughout aerobic bacterial growth, the p4070 promoter conferred stronger sfGFP expression than that conferred by pUP2n38, and weaker expression than that conferred by pRPLS **(Figure 17)**. In fact, sfGFP fluorescence reached 469 AUs for a culture OD_{600nm} of 1.162 AUs, which is equivalent to 56.2% of the expression conferred by the pRPLS_sfGFP construct, which contains its optimal *ldh* ribosomal binding site.

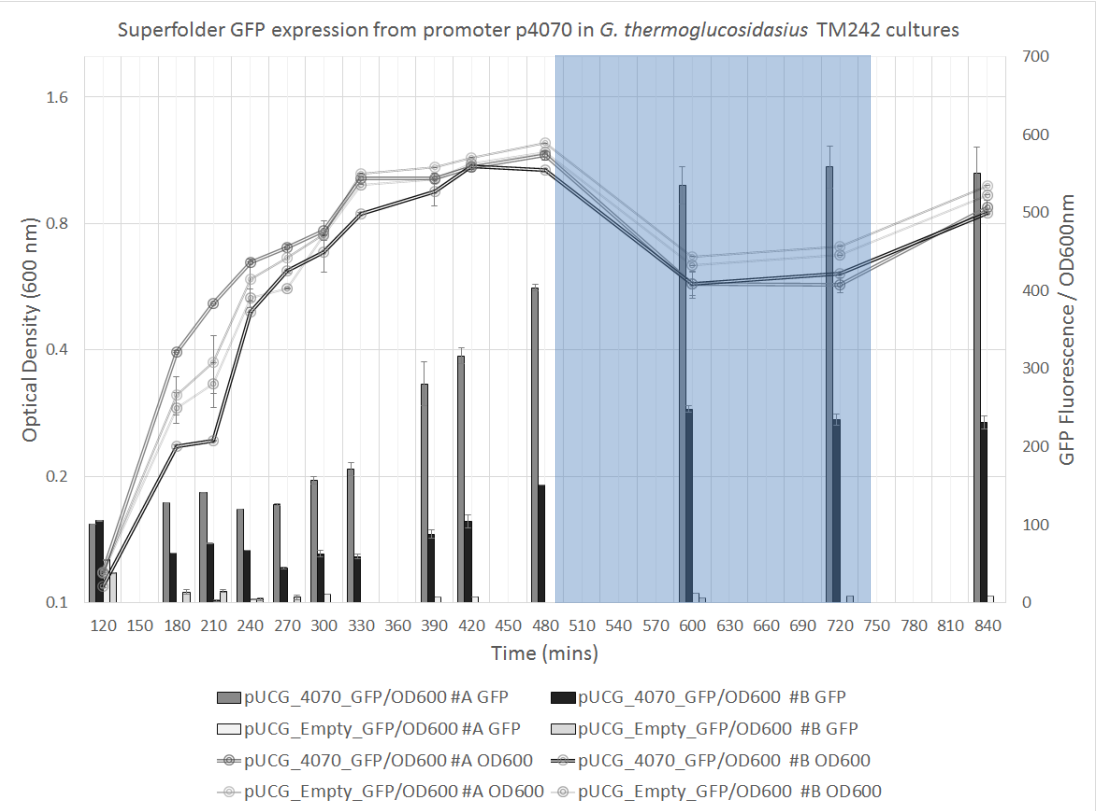


Figure 17 – Superfolder GFP expression profile from p4070 promoter. Superfolder GFP expression levels conferred by the p4070_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions (clear) in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Both biological duplicates and technical duplicates were measured. Once culture OD_{600nm} remained stable, an indicator of stationary phase, a transition towards oxygen-limiting conditions was achieved (shaded) by adding further media to reduce the closed culture air volume from 90% to 30%.

The p4070 promoter appears to have two intriguing properties. In contrast to the previous promoters, there are considerable differences in sfGFP expression between the duplicate cultures at a similar culture OD_{600nm}. Essentially, two duplicate cultures of the same p4070 expression constructs would have markedly different fluorescence profiles. However, their general expression profiles remain similar. Moreover, in contrast to the pRPLS and pUP2n38 promoters, increased sfGFP expression continues beyond the stabilisation of culture OD_{600nm} during stationary growth phase under aerobic conditions, and greater increases in sfGFP fluorescence are observed with an increase in culture OD_{600nm}. That is, that although the OD_{600nm} of the duplicate cultures was relatively constant between 330 and 480 minutes post-innoculation, the fluorescence readings continued to increase.

The intriguing expression profile of p4070 promoter adds further questions to the function of the encoded protein. One suggestion is that the function of the peg4070 product is the

utilisation of a surrounding carbon source into a transportable metabolite, given its weak homology to hydantoinases and histidinol dehydrogenases, which are involved in the production of N-carbomyl-amino acids and histidine, respectively. Nevertheless, the p4070 promoter may prove useful for the continuous expression of heterologous proteins in non-replicative cultures of *Geobacillus* spp.

3.5 The anaerobically-inducible pAdhE promoter.

One stratagem to circumvent the issue of increased metabolic burden incurred by simultaneous cellulase production and sugar fermentation is the upregulation of fermentation genes or antisense RNA-based downregulation of cellulase genes by exploiting anaerobically-induced promoters. The RNAseq analysis revealed several ORFs that showed significant upregulation under anaerobic conditions. As expected, the ORF with the highest upregulation under anaerobic conditions was that of Acetaldehyde Dehydrogenase (AdhE, peg3856) **(Table 7)**, which catalyses the conversion of acetaldehyde to ethanol during anaerobic fermentation (Extance *et al.*, 2013).

Table 7 – RNAseq data of *peg3856* ORF. Relative expression levels reported as the proportion of sequenced ORF transcripts (RPKM value) relative to total sequenced transcriptome (total RPKM). The average % change in RPKM values from aerobic duplicates to anaerobic duplicates calculated based on relative RPKM values, and is presented as a product of promoter strength (average relative RPKM under aerobic conditions).

		Relative Expression Levels			
		Aerobic 1	Aerobic 2	Anaerobic 1	Anaerobic 2
		0.004%	0.004%	1.347%	0.391%
		Average % change (Aerobic → Anaerobic)		**Average % change relative to promoter strength**	
peg3856	Acetaldehyde Dehydrogenase	21524.18%		0.87%	

Analysis of the region upstream of the *adhE* gene reveals the presence of strongly predicted -10 and -35 sites and a putative integration host factor (IHF) overlaying the -35 box **(Figure 18)**. IHF motifs have been reported in the anaerobic regulation of operons from other organisms, including the *E. coli pfl* operon (Sirko *et al.*, 1993).

Figure 18 – Upstream promoter region of *G. thermoglucosidasius* TM242 *adhE (peg3856)* gene. Promoter features were predicted using the BPROM promoter prediction server, and manually annotated using SnapGene Viewer.

To investigate its potential use as an anaerobically-inducible promoter in *G. thermoglucosidasius* TM242, the promoter region was amplified from *G thermoglucosidasius* TM242 genomic DNA using primers 12(F)/13(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method. As expected, no detectable sfGFP fluorescence was observed under aerobic growth conditions in 2xTY rich media **(Figure 19)**.

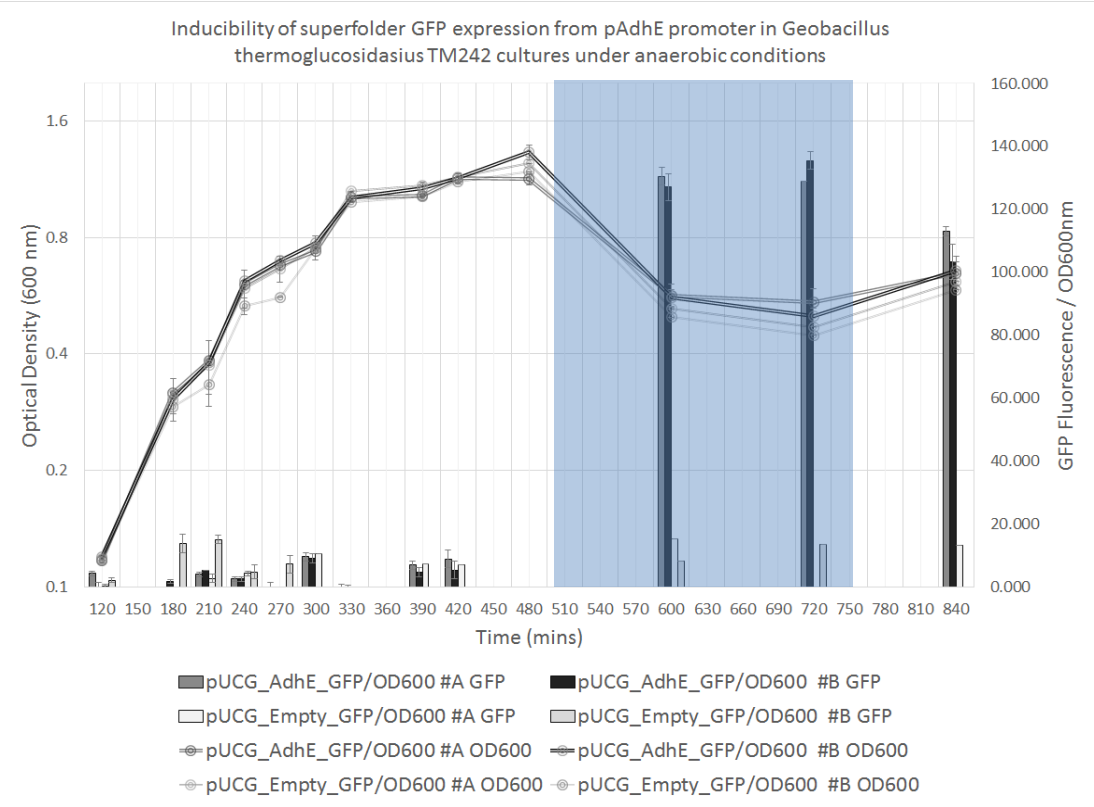


Figure 19 – Superfolder GFP expression profile from pAdhE promoter. Superfolder GFP expression levels conferred by the pAdhE_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions (clear) in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Both biological duplicates and technical duplicates were measured. Once culture OD_{600nm} remained stable, an indicator of stationary phase, a transition towards oxygen-limiting conditions was achieved (shaded) by adding further media to reduce the closed culture air volume from 90% to 30%.

Although the further addition of culture medium to achieve oxygen limiting conditions did initially dilute culture OD_{600nm} and sfGFP fluorescence by 83% (OD_{600nm} = 0.21, GFP = 0 (data not shown)), sfGFP expression levels rose to 69 ± 0 AUs at a measured culture OD_{600nm} of of 1.083 ± 0.069 AUs, a strong indication of promoter upregulation under oxygen limiting conditions.

Although induced promoter activity was weak, the low basal activity and strong inducibility of pAdhE may prove useful in expression studies of toxic or oxygen-sensitive proteins.

3.6 The strong and trehalose-inducible pTre promoter.

As shall be discussed in **section 4.1**, *G. thermoglucosidasius* TM242 can utilise trehalose as a sole carbon source, as has been reported for *Bacillus subtilis* 168 (Kennett and Sueoka, 1971). In *B. subtilis*, the utilisation of trehalose is dependent on the activity of three proteins that

make up the *trePAR* operon (Schöck and Dahl, 1996). Firstly, trehalose must be transported across the cell membrane and phosphorylated by the trehalose-specific PTS II system, encoded by *treP*. Trehalose-6-phosphate is subsequently hydrolysed by an intracellular phospho-α--glucosidase (*treA*) to β-D-glucose-6-phosphate, which feeds directly into the glycolysis I pathway, and D-glucopyranose, which is subsequently phosphorylated into D-glucopyranose-6-phosphate and also fed into the glycolysis I pathway. Expression of the *trePAR* operon has been reported to be regulated by the repressor TreR, which has been characterised to recognise and bind to two palindromes (trePO1 and trePO2) with the consensus sequence TGTATATACA (**Figure 20**) (Schöck and Dahl, 1996).

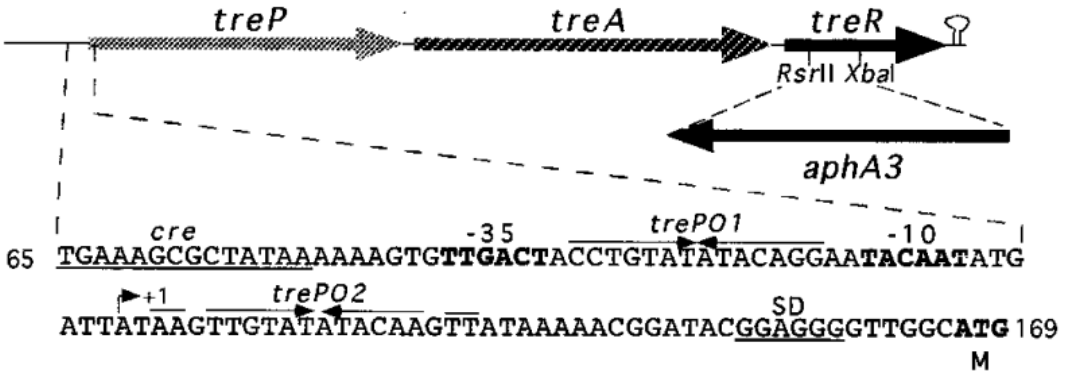


Figure 20 – *Bacillus subtilis trePAR* operon and characterised trehalose-inducible promoter (Schöck and Dahl, 1996).

Expression of the *trePAR* operon was shown to be influenced by growth phase dependence and carbon catabolite repression, but induction by the addition of trehalose in the growth medium was observed (Helfert *et al.*, 1995). Although inactivation of the central component of carbon catabolite repression, CcpA, did lead to a loss of TreA repression by fructose, repression by glucose was still present (Schöck and Dahl, 1996).

Analysis of the *G. thermoglucosidasius* TM242 genome reveals the presence of a similar *trePAR* operon, which is identical to that of *B. subtilis* 168 in both genotypic composition and orientation. Moreover, analysing the region upstream of *treP*, the first gene in the operon, reveals the presence of similar -10 and -35 sites and almost identical inverted repeats to trePO1 and trePO2, the characterised operator sequences of *B. subtilis* 168 TreR (**Figure 21**).

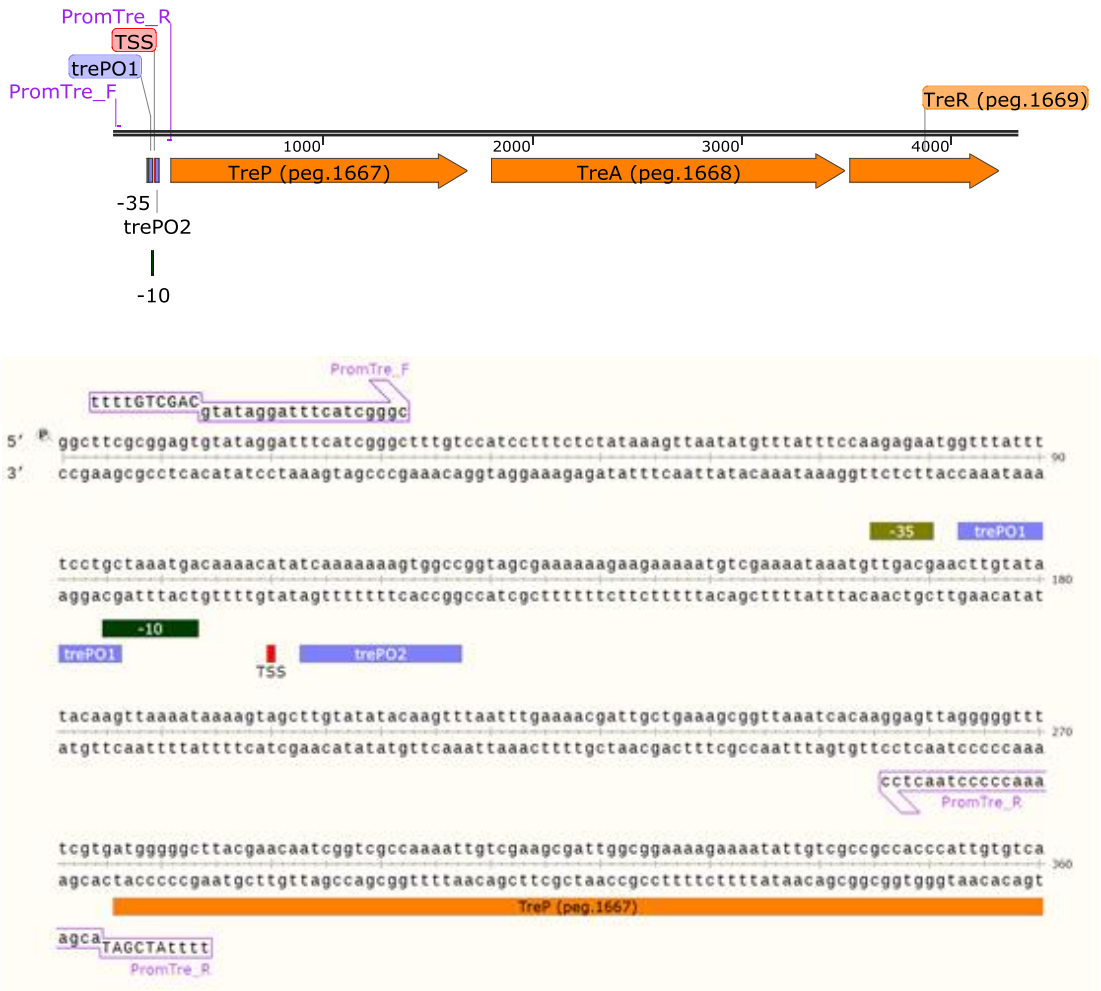


Figure 21 – *G. thermoglucosidasius* TM242 annotated *trePAR* operon and putative trehalose-inducible promoter region. Promoter features were predicted using the BPROM promoter prediction server, and manually annotated using SnapGene Viewer. Putative promoter operators trePO1 and trePO2 were identified manually.

To investigate its potential use as an inducible promoter in *G. thermoglucosidasius* TM242, the highlighted promoter region was amplified from *G. thermoglucosidasius* TM24 genomic DNA using primers 14(F)/15(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method. The basal expression of the construct under aerobic growth conditions was investigated in 2xTY rich media. 2xTY rich media contains yeast extract, which is likely to contain trehalose phosphate, a component of yeast cell walls. Although promoter activity under these conditions was expected, the considerable strength of the pTre promoter was unexpected. For example, GFP fluorescence by *G. thermoglucosidasius* TM242 pUCG_pTre_sfGFP cultures reached 500 AU at a measured culture OD_{600nm} of 1 AU, a fluorescence value only 75% of that observed in *G. thermoglucosidasius* TM242 pUCGRPLS_sfGFP cultures (700 AU) **(Figure 22)**. Remarkably, the

optimisation of the pTre promoter RBS, which resulted in a 2.5-fold increase in pRPLS promoter strength, is likely to increase expression levels higher than those observed with the optimised pRPLS_BstRBS expression fragment. Unlike the pRPLS and pUP2n38 promoters, which conferred static sfGFP expression at stationary growth phase, an increase in sfGFP expression by the pTre promoter continues during stationary growth phase.

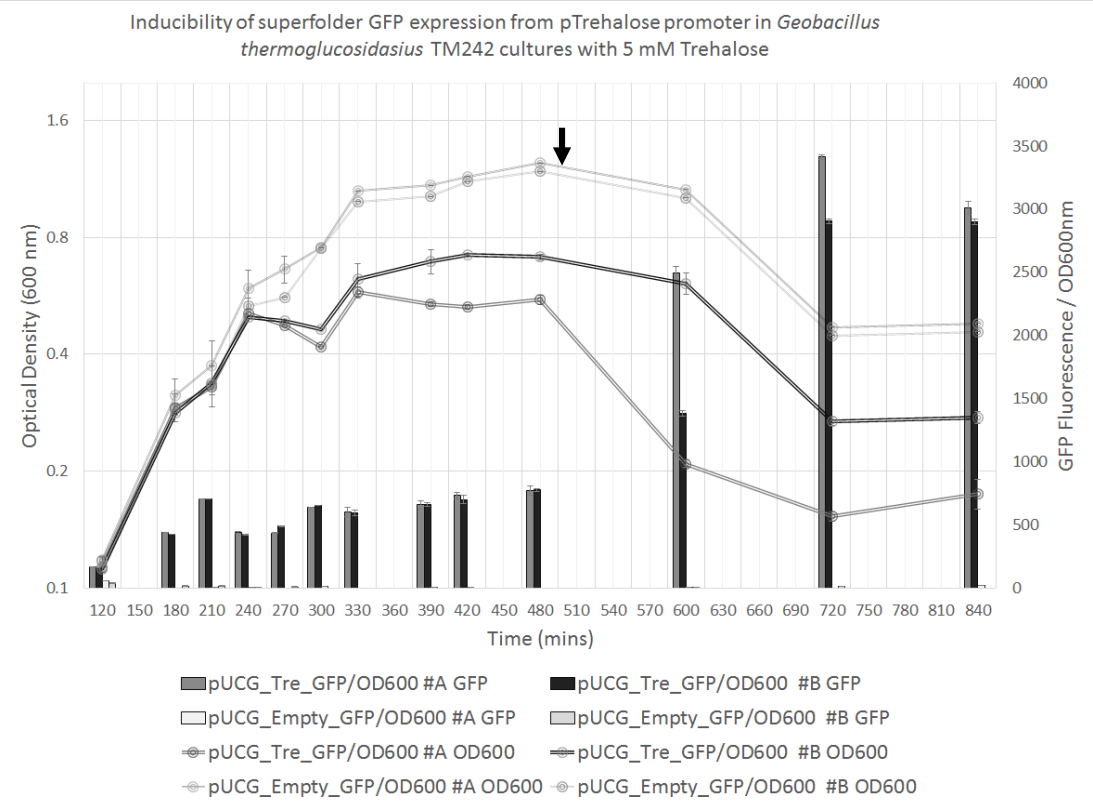


Figure 22 – Superfolder GFP expression profile from pTre promoter. Superfolder GFP expression levels conferred by the pTre_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Once culture OD_{600nm} remained stable, an indicator of stationary phase, 5 mM Trehalose was added (thick arrow). Both biological duplicates and technical duplicates were measured.

The strong expression conferred by the pTre promoter, with further inducibility by trehalose addition, makes it an intriguing promoter to further characterise.

3.7 The sorbitol-inducible pSorb promoter.

G. thermoglucosidasius TM242 has been observed to utilise sorbitol (glucitol) as a sole carbon source, but not its stereoisomer, galactitol (dulcitol) **(Section 4.1)**. Therefore, the presence of a putatitive galactitol utilisation operon on the chromosome of *G. thermoglucosidasius* TM242 was surprising **(Table 8)**. Analysis of the region upstream of the first ORF gene (*galR*)

reveals the presence of strongly predicted -10 and -35 sites and a two inverted repeats, the first of which overlaps the -10 box, suggesting transcriptional repression by transcription factor binding over the -10 box.

Gene	AA Length	Predicted gene product
peg1126	2079	Predicted galactitol operon regulator
peg1127	468	PTS system, galactitol-specific IIA component (EC 2.7.1.69)
peg1128	285	PTS system, galactitol-specific IIB component (EC 2.7.1.69)
peg1129	1266	PTS system, galactitol-specific IIC component (EC 2.7.1.69)
peg1130	1032	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)
peg1131	195	hypothetical protein
peg1132	1053	Galactitol-1-phosphate 5-dehydrogenase (EC 1.1.1.251)

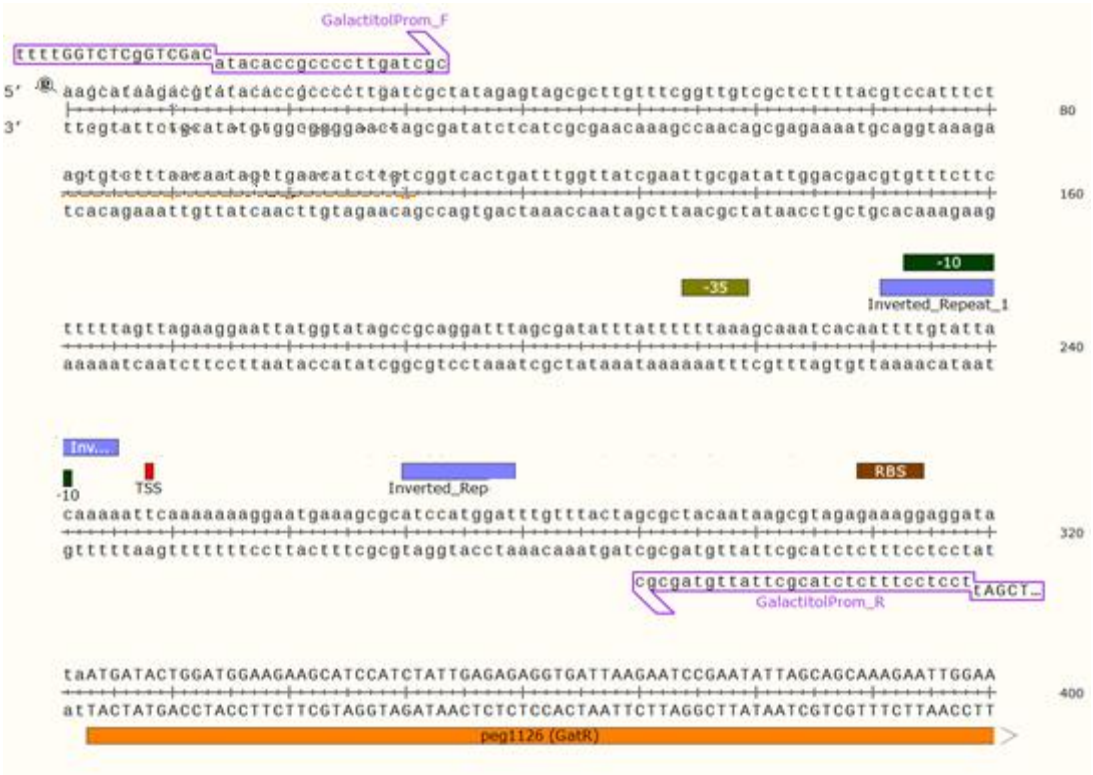


Table 8 and Figure 23 – *G. thermoglucosidasius* TM242 annotated *gal* operon and putative galactitol/sorbitol-inducible promoter region. Promoter features were predicted using the BPROM promoter prediction server, and manually annotated using SnapGene Viewer. Inverted Repeats 1 and 2 were identified using PALINDROME (Pasteur Institute).

To investigate its potential use as an inducible promoter in *G. thermoglucosidasius* TM242, the highlighted promoter region was amplified from *G. thermoglucosidasius* TM242 genomic DNA using primers 16(F)/17(R) and inserted upstream of the superfolder GFP gene using the aforementioned GoldenGate method. As expected, only very low levels of sfGFP fluorescence (10 AUs) were observed under aerobic growth conditions in the absence of sorbitol.

The addition of 5 mM galactitol to similar duplicate cultures resulted in no detectable increase in sfGFP expression (data not shown). However, with the addition of 5 mM sorbitol, detectable GFP expression increased to a fluorescence of 80 AU **(Figure 24)**. Although unexpected, this result may indicate that a downstream metabolic intermediate of sorbitol and galactitol, such as *keto*-D-fructose or β-D-fructofuranose, is the likely ligand for transcriptional derepression. Moreover, in contrast to the 5 mM trehalose induction of pUCG_pTre_sfGFP constructs, the sorbitol induced cultures continued to grow further to a culture OD_{600nm} of 1.92 AUs, which was expected given its observed sorbitol utilisation capabilities.

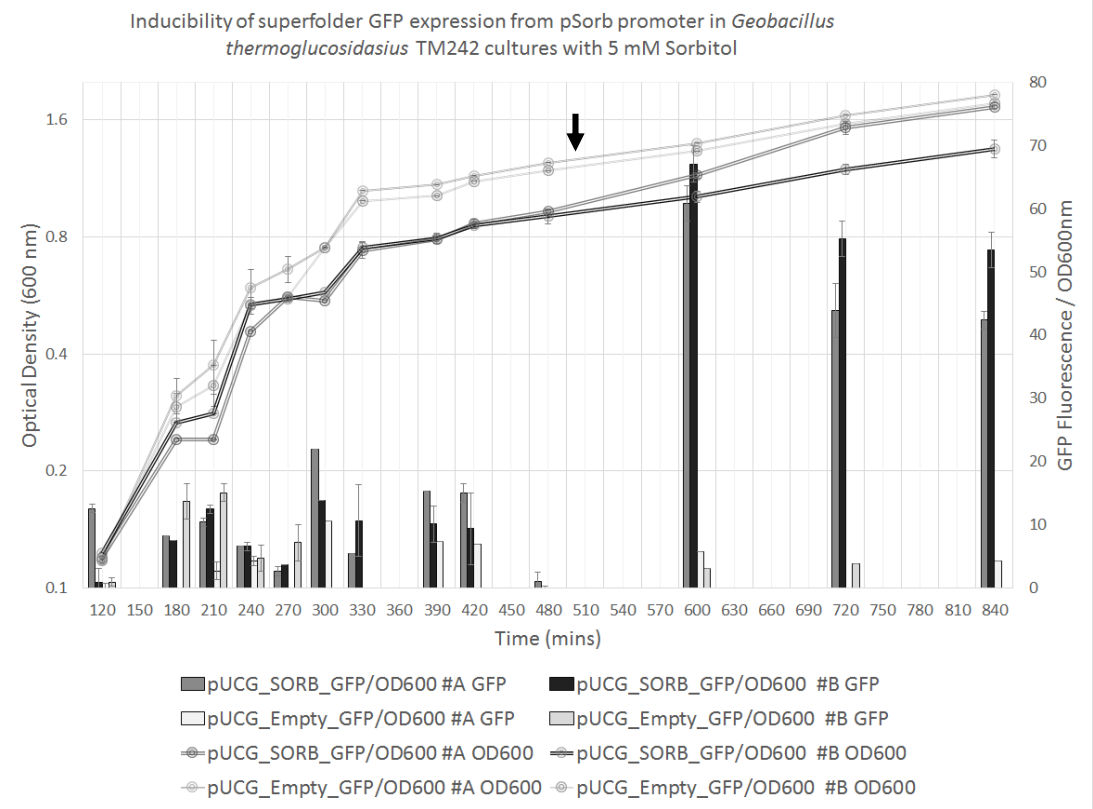


Figure 24 – Superfolder GFP expression profile from pSorb promoter. Superfolder GFP expression levels conferred by the pSorb_sfGFP construct in *G. thermoglucosidasius* TM242 during aerobic growth conditions in 10 ml rich 2xTY media cultures in 50 ml centrifugation tubes. Once culture OD_{600nm} remained stable, an indicator of stationary phase, 5 mM Sorbitol was added (thick arrow). Both biological duplicates and technical duplicates were measured.

Although inducible expression conferred by the pSorb promoter was weak, the strong inducibility of the promoter by sorbitol addition makes it a powerful tool for transient expression of heterologous proteins in *Geobacillus* spp. However, further characterisation of

the inducibility of the promoter is required with different sugar alcohols at varying concentrations.

In conclusion, the studies of native and synthetic expression fragments is of paramount importance for the future development of *G. thermoglucosidasius* as a chassis organism. The constitutive RPLS promoter will serve well as the core module for the high-level expression of glycosyl hydrolases, whereas the weaker pUP2n38 promoter will be utilised for ancillary proteins or proteins with toxicity in high concentrations. The characterisation of native *G. thermoglucosidasius* promoters has provided exciting possibilities for the performance of ligand-specific inducible expression studies, as well as further characterisation of the mysterious *peg4070* gene.

Chapter 4: The investigation of *G. thermoglucosidasius* catabolic capabilities and prospection of glycosyl hydrolases

4.1 Genomic analysis of *G. thermoglucosidasius* TM242 lignocellulolytic capabilities

By approaching CBP engineering holistically, the effective degradation of lignocellulosic biomass by an engineered *G. thermoglucosidasius* strain is based purely on the genetically-encoded capabilities of the strain to express the required glycosyl hydrolases, sugar transportation systems, and downstream catabolic enzymes required to break down a specific polysaccharide, transport the shorter chain oligo-, di- or monosaccharides into the cytoplasm, and subsequently introduce them into the central metabolism of the cell.

As mentioned on **Section 1.6**, the *G. thermoglucosidasius* NCIMB11955 wild-type strain, which is the parent strain of the engineered ethanologenic TM242 strain, is capable of metabolising and fermenting C5 sugars, such as D-xylose and L-arabinose, and C6 sugars, such as glucose and galactose (Cripps *et al.*, 2009). This is advantageous for a candidate strain chosen for CBP engineering, since it can exploit the wide range of sugars released from lignocellulosic biomass degradation.

Fermentative capabilities of *G. thermoglucosidasius* TM242

To confirm these metabolic capabilities in the ethanologenic *G. thermoglucosidasius* TM242 strain, the API CHB 50 test (*bioMérieux*) was used to measure the strain's fermentative capabilities on 50 carbon sources **(Table 9)**. *G. thermoglucosidasius* TM242 was capable of fermenting 22 out of the 50 carbon sources, including the major monosaccharides and disaccharides released from hydrolysis of *Miscanthus* x *giganteus* (D-glucose, D-xylose, L-arabinose and cellobiose).

Xylobiose, the second major disaccharide released from hydrolysis of *Miscanthus* x *giganteus*, was not available in the API CHB 50 test. However, a *xynB2* gene is present in the genome, which encodes a 1,4-β-xylosidase that is likely to facilitate the cleavage of xylobiose to D-xylose (Espina *et al.*, 2014), which the strain can metabolise.

Table 9 – API 50 CH analysis of the fermentation capabilities of *G. thermoglucosidasius* TM242 on multiple sugars. *G. thermoglucosidasius* TM242 colonies were transferred from 2TY rich agar to the provided dispersing medium, and tested in duplicates according to the API 50 manual (bioMérieux, Inc.). Positive (+), negative (-), and variable (Var.) results were observed. Analysis performed by Chris Vennard (University of Bath) and Carolyn Williamson (University of Bath).

Carbon Source	TM242 Ferm.	Carbon Source	TM242 Ferm.	Carbon Source	TM242 Ferm.
Glycerol	Var	Inositol	-	Inulin	-
Erythrol	-	Mannitol	+	D-melezitose	-
D-arabinose	-	Sorbitol	+	D-raffnose	-
L-arabinose	+	α-methyl-D-mannoside	-	Amidon	+
Ribose	+	α-methyl-D-glucoside	+	Glycogen	-
D-xylose	+	N-acetyl-glucosamine	+	Xylitol	-
L-xylose	-	Amygladin	+	Gentiobiose	+
Adonitol	-	Arbutin	+	D-turanose	+
B-methyl-D-xyloside	-	Esculin	+	D-lyxose	-
Galactose	-	Salicin	+	D-tagatose	-
D-Glucose	+	Cellobiose	+	D-fucose	-
Fructose	+	Maltose	+	L-fucose	-
Mannose	+	Lactose	-	D-arabitol	-
L-sorbose	-	Mellibiose	-	L-arabitol	-
Rhamnose	-	Saccharose	+	Potassium gluconate	-
Galactitol	-	Trehalose	+	2-ketogluconate	-

The Hemicellulose Utilisation Locus of *G. thermoglucosidasius* TM242

Generally, *Geobacillus* strains have been shown to be highly effective in the degradation of hemicellulose polymers, and the molecular determinants underlying the capacity of this genus to utilize hemicellulose have been shown to reside on a single genetic island (De Maayer *et al.*, 2014). A comparison study of this hemicellulose utilisation (HUS) locus between 23 *Geobacillus* strains revealed extensive genetic variability among the strains, despite the island being localized in a common genomic position in most strains **(Figure 25)** (De Maayer *et al.*, 2014). Four of the compared strains in this study were *G. thermoglucosidasius* strains (C56-YS93, Y4.1MC1, M10EXG and TNO-09.20), but no analysis on the HUS locus of the subject *G. thermoglucosidasius* TM242 strain has been carried out.

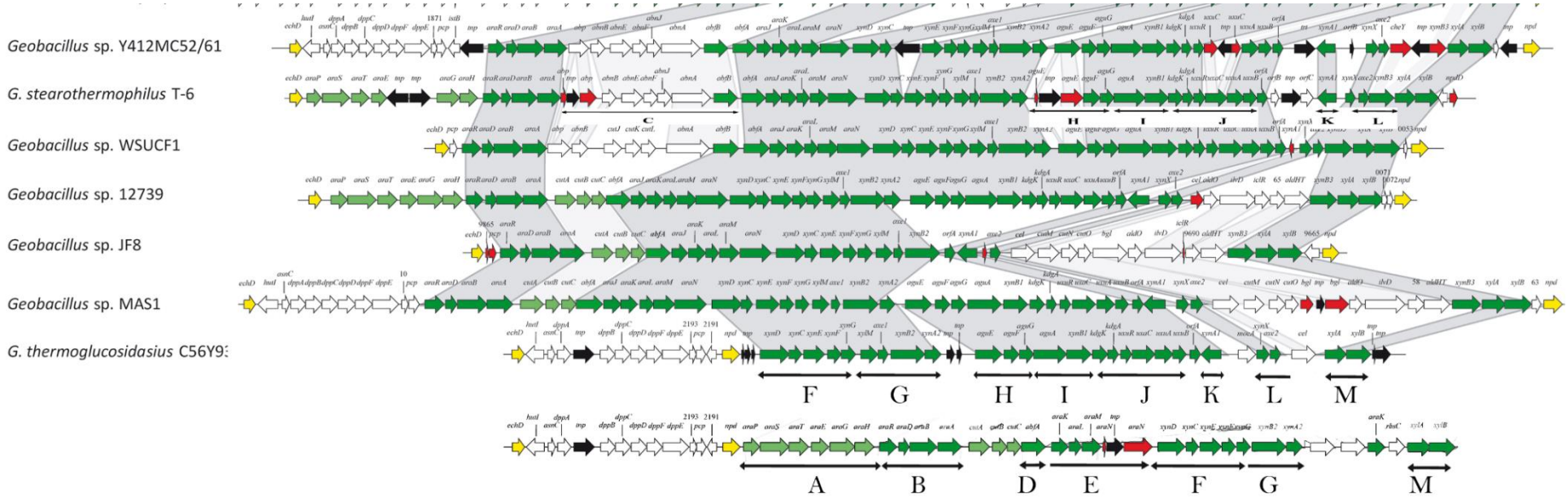
To this end, a HUS island was identified on the genome of *G. thermoglucosidasius* TM242 and added to the HUS locus mapping performed by De Maayer and colleagues. Similar to *G. thermoglucosidasius* C56-YS93, the locus starts with a gene cluster (dppABCDFE) that encodes a predicted oligopeptide transporter. However, in contrast to C56-YS93, the TM242 strain has the encoded capabilities for intracellular arabinofuranose degradation (Gene Cluster D), associated transport functions (Gene Cluster A) and metabolism of L-arabinose (Gene Cluster B). Intriguingly, the strain has lost Gene Cluster C, which encodes a set of seven extracellular arabinan degrading enzymes (arabinases and arabinofuranosidases), which may point towards the strain's scavenging capabilities. However, it encodes an alternative pathway for pentose sugar metabolism (Gene Cluster E) which is absent from *G. thermoglucosidasius* C56-YS93.

The utilisation of xylan is more synonymous between the two strains, with both strains encoding the capabilities for trans-membrane transport of (Gene Cluster F) and subsequent intracellular degradation of xylooligosaccharides (Gene Cluster G and L), followed by metabolism of the xylose monosaccharide (Gene Cluster M). However, *G. thermoglucosidasius* C56-YS93 encodes three gene clusters involved in aldotetraouronic acid (Gene Cluster H), aldotetraouronic acid (Gene Cluster I) and glucuronic acid (Gene Cluster J) utilisation, which are absent from TM242. Another unique distinction of the *G. thermoglucosidasius* TM242 HUS locus is the presence of a four open reading frames between gene clusters G and M, with one ORF predicted to be a β-xylosidase.

One explanation for the significant variation in HUS loci between two strains of the same species is the possibility that, in their natural environments, *Geobacillus* strains form

consortia, whereby strains deficient in genes or pathways required for hemicellulose degradation and utilization may be complemented by other members of the consortia. However, it could be argued that this level of complementation is unlikely given the limited number of extracellular hydrolytic enzymes produced by *Geobacillus*.

Another explanation for the variability among *Geobacillus* strains in regards to their hemicellulose utilisation capabilities is that they have adapted to degrade distinct hemicellulose substrates, such as are found in different plant species and tissues. The lack of genes encoding an α-glucuronidase (*aguA*), its associated transport system (*aguEFG*) and uronate metabolic genes in *Geobacillus thermoglucosidasius* TM242 implies that this strain utilises xylo-oligosaccharides which are not substituted with glucuronic acids, while a xylan polymer devoid of arabinofuranosyl substituents is the likely target of *G. thermoglucosidasius* C56YS93. Genes coding for ComG and ComK competence factors are present in *Geobacillus* spp. genomes, and, if functionally expressed, suggest that a degree of horizontal gene transfer may be feasible for the hypothesised adaptation of the HUS islands to occur.



Cluster	Function
A	L-Arabinose transport
B	Arabinose metabolism
C	Arabinan utilization
D	Intracellular arabinofuranose degradation
E	Alternative pathway for pentose sugar metabolism

Cluster	
F	Xylooligosaccharide transport
G	Intracellular xylooligosaccharide degradation
H	Aldotetraouronic acid transport
I	Intracellular Aadotetraouronic acid degradation
J	Glucuronic acid metabolism

Cluster	
K	Extracellular xylan degradation
L	Intracellular xylooligosaccharide degradation
M	Xylose metabolism

Figure 25 - Comparative diagram of the *Geobacillus* spp. HUS loci (adapted from (De Maayer *et al.*, 2014). Aligned HUS islands of the eight sequenced *Geobacillus* spp. genomes, including that of *G. thermoglucosidasius* TM242. In contrast to the conventional arrangement of flanking *echD* and *npd* genes, coloured in yellow, the HUS loci of *G. thermoglucosidasius* TM242 and C56-YS93 contain an *echD-npd* flanked operon upstream of the hemicellulose utilisation genes. Genes coding for transposons are coloured in black, and genes in which the reading frames are transposon-disrupted are coloured in red. Genes conserved among >70% of 20 HUS+ strains are coloured in green. Those conserved among >50% and <70% in light green, and those conserved among <50% of the HUS+ strains are coloured in white.

4.2 Prospecting for component glycosyl hydrolases for the recombinant system

Since *G. thermoglucosidasius* NCIMB 11955 and its mutants can metabolise the monosaccharide and disaccharide components of cellulose and xylan, the major components of *Miscanthus* x *giganteus*, the fundamental requirement for engineering of a CBP strain is the heterologous expression of highly active glycosyl hydrolases to liberate these metabolisable sugars from the polysaccharides.

A variety of novel glycosyl hydrolases are regularly being reported and characterised, but for the purpose of this study, the focus will be on the main saccharolytic activities of exoglucanases (EC 3.2.1.91/3.2.1.176), endo-β-1,4-glucanases (EC 3.2.1.4), endo-β-1,4-xylanases (EC 3.2.1.8) and endo-β-1,3-xylanases (EC 3.2.1.32).

The thermophilic nature of *Geobacillus spp.* leads to the prerequisite for both thermostability and thermoactivity of these heterologous glycosyl hydrolases at temperatures of at least 50°C-65°C, the range at which *G. thermoglucosidasius* is conventionally cultured. Therefore, the search for potential glycosyl hydrolase candidates was focused on a well-studied panel of thermophilic cellulolytic organisms; namely *Clostridium thermocellum* DSM1237*, Caldicellulosiruptor saccharolyticus* DSM8903*, Clostridium clariflavum* DSM19732*, Thermobifida fusca* XY and *Thermotoga maritima* DSM3109.

Glycosyl hydrolases are generally characterised by their specific activities on polysaccharide assay substrates, including amorphous carboxymethylcellulose (CMC), crystalline cellulose Avicel, phosphoric acid swollen cellulose (PASC) and barley β–glucan (BBG). By searching the available peer-reviewed literature on characterised glycosyl hydrolases, a database of thermoactive glycosyl hydrolases with their specific activities on the aforementioned substrates was compiled **(Figure 30 and APPENDIX II)**.

Exoglucanases

Due to the greater recalcitrance and carbon-density of crystalline cellulose, there was a greater emphasis on the expression and secretion of crystalline-cellulose degrading exoglucanases. In fact, it has been observed that endoglucanase activity on complex cellulosic biomass, like *Miscanthus* x *giganteus*, is markedly improved by the release of amorphous cellulose from crystalline cellulose fibres through exoglucanase activity (Banerjee *et al.*, 2010). This is one of four types of synergism demonstrated in cellulase systems: synergism between endoglucanases and exoglucanases, synergism between reducing- and nonreducing-end exoglucanases, synergism between processive endoglucanases and both

endoglucanases and exoglucanases, and synergism between β-glucosidases and other cellulases (Bhat and Bhat, 1997).

Table 10 – Biochemical properties and specific activities of characterised bacterial exoglucanases. Optimal temperatures (T_{opt} in °C), optimal pH, and specific activities (U/mg) of exoglucanases reported on multiple cellulosic substrates.
* enzyme activity characterised in the absence of CBM3-containing scaffoldin protein.

Identifier	Organism	T_{opt}	pH_{opt}	CMC	Avicel	PASC	βBG	Ref.
TFcel48A	*T. fusca* XY	50	6.5	0.292	-	0.405	-	(Irwin *et al.*, 2000)
TFcel6B	*T. fusca* XY	60	7.0	0.17	0.25	-	15.8	(Calza *et al.*, 1985)
CTcel48S*	*C. thermocellum* DSM1237	70	5.5	0.008	0.019	0.15	-	(Kruus et al., 1995b)
CTcel48Y	*C. thermocellum* DSM1237	65	7.0	0.0017	0.0025	0.299	0.158	(Berger *et al.*, 2007)

Moreover, in a study to optimise the enzyme mixture required for glucose release from *Miscanthus* x *giganteus*, the optimal exoglucanase proportion was 42%, 48% and 36% for the degradation of AFEX-pretreated, alkaline-peroxidase pretreated, and 0.25 M NaOH-pretreated *Miscanthus* x *giganteus,* respectively (Banerjee *et al.*, 2010).

However, in contrast to endoglucanases, there are few characterised bacterial exoglucanases reported with notable activities at elevated temperatures. Nevertheless, five exoglucanases of notable activity, and thermostable at *Geobacillus* growth temperatures, have been published with specific activities determined on the crystalline-cellulose commercial substrate Avicel **(Table 10)**.

***C. thermocellum* Cel48S and Cel48Y**

C. thermocellum DSM1237 expresses two glycosyl hydrolase family 48 reducing-end exoglucanases of interest.

The cellulosomal *Ct*cel48S (Cthe_2089) is the most abundant subunit of cellulosomes, and has been characterised to degrade crystalline cellulose synergistically with the anchor protein CipA (Kruus et al., 1995b, Wang et al., 1994). This is most likely due to the absence of any CBM domains for exoglucanase attachment to the cellulose substrate, which is a common feature of processive non-cellulosomal exoglucanases. The presence of two dockerin domains directs attachment of *Ct*cel48S to the cohesin domains of the anchoring protein CipA, which contains a CMB3a domain, thereby facilitates a degree of processivity for the attached exoglucanase (Kruus et al., 1995b). It is therefore not surprising that recombinant *Ct*cel48S, expressed and purified in *E. coli* and characterised in the absence of CipA, has been reported to exhibit extremely low specific activity against crystalline Avicel. However, it exhibited much higher activity on amorphous PASC, further providing evidence of the importance of CBM3a domains in effective cellulase activity on crystalline substrates.

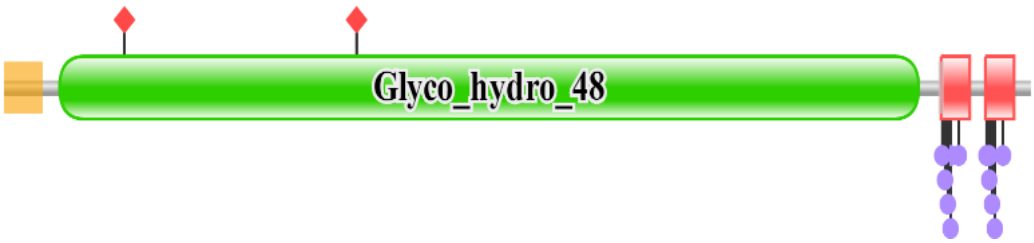


Figure 26 – Annotated protein domain structure of cellulosomal *C. thermocellum* exoglucanase S, consisting of an N-terminal signal peptide (yellow box), glycosyl hydrolase family 48 catalytic domain and dual dockerin domains (red boxes). Active site residues are marked as red diamonds, and metal binding residues are marked by purple circles.

The major hydrolysis product from amorphous cellulose and Avicel degradation is cellobiose, although cellotriose was also detected at an approximate ratio of 5:1 cellobiose/cellotriose (Kruus et al., 1995b). Moreover, cellopentaose was hydrolysed to cellobiose and cellotriose in an equimolar ratio, suggesting that *Ct*cel48S is not able to hydrolyse cellotriose. This may not be problematic since *G. thermoglucosidasius* has been shown to metabolise cellulo-oligosaccharides longer than cellotriose (*Rebio*, personal communication).

In contrast, *Ct*cel48Y (Cthe_0071), a non-cellulosomal exoglucanase, does contain a CBM3b module at its C-terminus and does not contain a dockerin module. In fact, since the sequence of the catalytic GH48 module is most closely related (59% identity) to the catalytic GH48 module of the aforementioned *Cs*cel48/9A, it has been noted that the distant relationship

between *Ct*cel48S and *Ct*cel48Y could be an indication of a different source of both genes (Berger *et al.*, 2007).

Figure 27 - Annotated protein domain structure of non-cellulosomal *C. thermocellum* exoglucanase Y, consisting of an N-terminal signal peptide (red box), glycosyl hydrolase family 48 catalytic domain and carbohydrate binding module family 3 domain (red oval). Active site residues are marked as purple diamonds, and unstructured regions are marked by grey boxes.

*Ct*cel48Y has been shown to degrade amorphous PASC more effectively than barley β-glucan and carboxymethylcellulose. However, the main point of interest is the characterised synergistic activity of *Ct*cel48Y with the soluble processive endoglucanase *Ct*Cel9I (Berger *et al.*, 2007). The simultaneous presence of *Ct*cel48Y and *Ct*cel9I resulted in a 2.1-fold higher activity than that of the summed activity of the individual enzymes (Berger *et al.*, 2007). The degree of synergism was characterised to be highest at an approximate ratio of 17:1 *Ct*cel48Y:*Ct*cel9I (Berger *et al.*, 2007). Therefore, these two non-cellulosomal cellulases are thought to possibly code for a second, soluble true cellulase system in *C. thermocellum* DSM1237.

Thermobifida fusca Cel48A and Cel6B

A second set of exoglucanases were identified from the Gram-positive thermophilic bacterium *Thermobifida fusca* YX. In fact, these two exoglucanses are natively produced at levels roughly four times greater than that of the native endocellulases (Irwin *et al.*, 2000).

Similar to other GH48 exoglucanases, *Tf*cel48A (E6, *Tfu_1959*) attacks the reducing end of cellulose fibres (Irwin *et al.*, 2000). When expressed, *Tf*cel48A has been reported to be expressed at 134 $mg \cdot L^{-1}$ or 34% of total extracellular cellulases produced (Spiridonov and Wilson, 1998), highlighting its natural importance in carbon source utilisation by *T. fusca*.

The major hydrolysis product from PASC degradation is cellobiose, with only a very small trace of cellotriose, which it does not cleave (Irwin *et al.*, 2000). Although *Tf*cel48A activity on various substrates is extremely low, synergistic filter paper assays with other enzymes from *T. fusca* resulted in increases in sugar release far greater than the additive activities of the individual enzymes (Irwin *et al.*, 2000) .

Figure 28 - Annotated protein domain structure of *Thermobifida fusca* reducing-end exoglucanase 48A, consisting of an N-terminal signal peptide (yellow box), carbohydrate binding module family 2 domain and glycosyl hydrolase family 48 catalytic domain. Active site residues are marked as purple diamonds, and unstructured regions are marked by grey boxes.

Table 11 – Synergistic filter paper assays with *T. fusca* cellulases. (Irwin *et al.*, 2000).

Enzyme mixture	Molar ratio used (in order as labelled)	Activity (U/µmol enzymes)
***Tf*cel5A (E5) endo**		0.788
***Tf*cel6B (E3) NR exo[b]**		0.222
***Tf*cel48A (E6) R exo[c]**		0.084
***Tf*cel6B + *Tf*cel48A**	1 : 1	0.522
***Tf*cel6B + *Tf*cel5A**	4 : 1	4.11
***Tf*cel6B + *Tf*cel5A + *Tf*cel48A**	2 : 1 : 2	7.02

Glycosylation of *Tf*cel48A is minimal, with only a very faint glycosylation band revealed by DIG glycan detection analysis (Irwin *et al.*, 2000).

In contrast, the same analysis with *T. fusca* Cel6B revealed very dark protein bands (Irwin *et al.*, 2000). *Tf*cel6B is a member of the glycosyl hydrolase 6 family of non-reducing end exoglucanases, and has activity similar to that of the fungal *T. reesei* exocellulase Cel6A, although *Tf*Cel6B has higher thermostability and a much broader pH optimum (Zhang *et al.*, 1995). Moreover, analysis of the *Tf*cel6B crystal structure argues that its long substrate tunnel makes the enzyme resemble the more powerful GH7 fungal cellobiohydrolases, and further suggests that *Tf*cel6B *is* more processive and has lower endoglucanolytic activity than the fungal GH6 exoglucanases (Sandgren *et al.*, 2013). Although *Tf*Cel6B exhibits weak activity on all cellulosic substrates, its merit is the synergism it exhibits with *Tf*cel48A and endoglucanases **(Table 11)**.

Figure 29 - Annotated protein domain structure of *Thermobifida fusca* non-reducing-end exoglucanase 6B, consisting of an N-terminal signal peptide (yellow box), carbohydrate binding module family 2 domain and glycosyl hydrolase family 6 catalytic domain. Active site residues are marked as purple diamonds, metal binding residues are marked by blue circles, and unstructured regions are marked by grey boxes.

To investigate *T. fusca* exoglucanase glycosylation further, the amino acid sequence of *Tf*Cel6B and *Tf*Cel48A was analysed using the N-glycosylation prediction server NetNGluc (*Technical University Denmark*)**.** As expected, analysis of the *Tf*cel48A revealed no obvious glycosylation sites in its mature protein-coding sequence. However, the sequence of *Tf*cel6B contained 7 putative N-glycosylation sites, which could prove to be problematic for heterologous expression if there is no conventional *N*-glycosylation machinery in *G. thermoglucosidasius*, as the KEGG pathway database appears to suggest. Nevertheless, if successful, the heterologous expression of functional *Tf*cel6B may provide some interesting preliminary data on the putative glycosylation capabilities of *G. thermoglucosidasius* TM242.

To facilitate the heterologous expression of these exoglucanases in *G. thermoglucosidasius* TM242, primer pairs were designed to amplify the mature protein coding sequences (MCS) of *TF*cel48F (primer pairs 18(F)/19(R), 20(F)/21(R), 22(F)/23(R)), *TF*cel6B (primer pair 24(F)/25(R)), *Ct*cel48S (primer pair 26(F)/27(R)), *Ct*cel48Y (primer pair 28(F)/29(R)), and *Cs*cel48/9A (primer pairs 30(F)/31(R), 32(F)/33(R), 34(F)/35(R)).

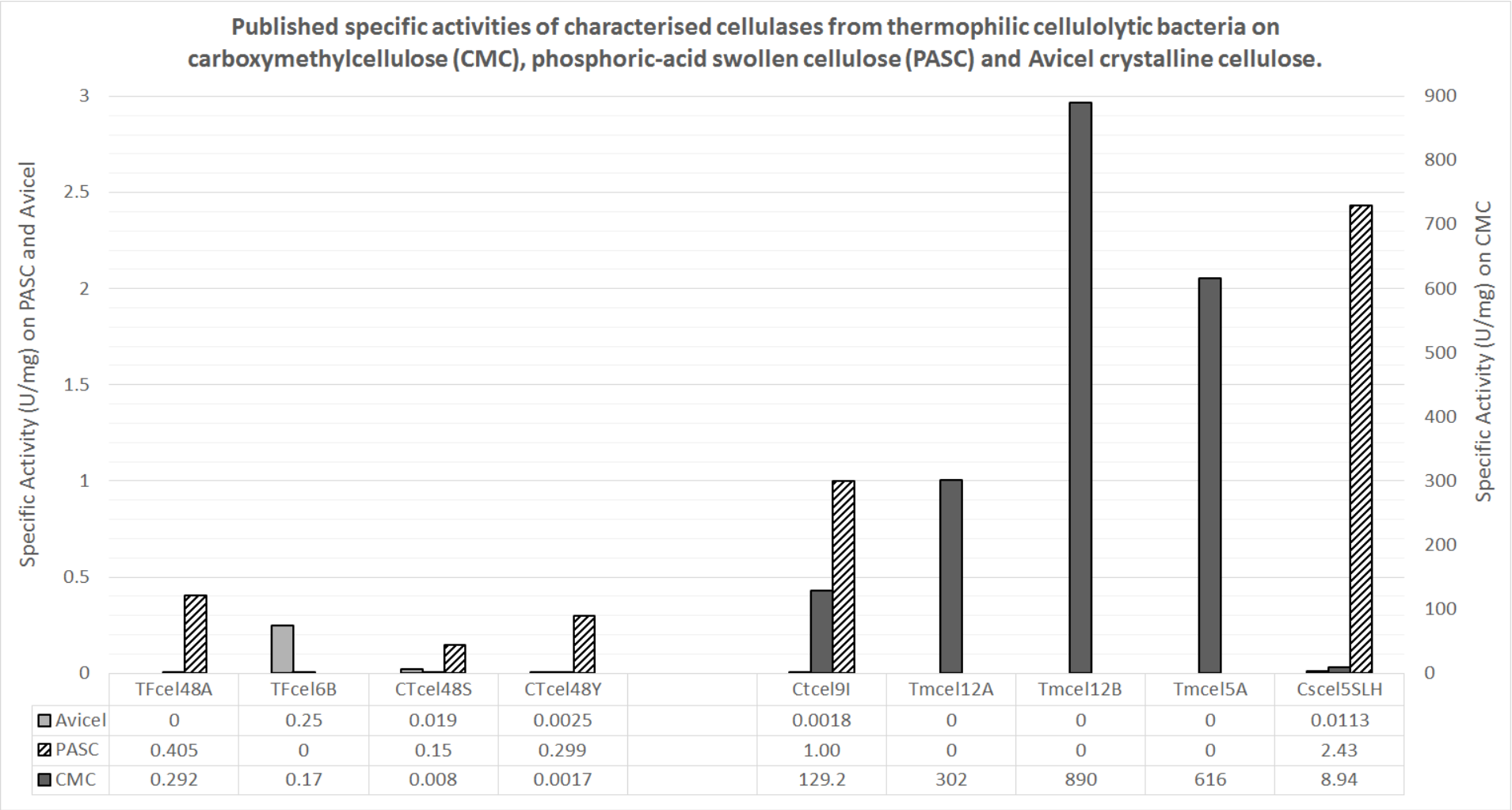


Figure 30 - Published specific activities of glycosyl hydrolases naturally expressed in thermophillic cellulolytic bacteria. ** denotes data derived by conversion of data from U/µmol unit to U/mg unit. NR* denotes specific activities that were not reported by the corresponding characterisation study.

Endoglucanases

Due to the lower recalcitrance of amorphous cellulose, the specific activities of endoglucanases assayed on amorphous CMC are generally greater than those of exoglucanases assayed on crystalline Avicel.

Table 12 – Biochemical properties and specific activities of characterised bacterial endoglucanases. Optimal temperatures (T_{opt}) in °C, optimal pH, and specific activities (U/mg) of endoglucanases reported on CMC (carboxymethylcellulose), microcrystalline Avicel, PASC (phosphoric acid swollen cellulose) and βBG (β-Barley Glucan).

Identifier	Organism	T_{Opt}	pH_{Opt}	CMC	Avicel	PASC	βBG	Reference
***Ct*cel9I**	*C. thermocellum* DSM1237	70	5.5	129.2	0.18	1.00	-	(Gilad *et al.*, 2003)
***Tm*cel12A**	*T. maritima* DSM3109	90	7	302	0	-	1785	(Bronnen meier *et al.*, 1995)
***Tm*cel12B**	*T. maritima* DSM3109	80	6	890	0	-	2905	(Bronnen meier *et al.*, 1995)
***Tm*cel5A**	*T. maritima* DSM3109	80	6	616	-	-	2345	(Pereira *et al.*, 2010)
***Cs*cel5SLH**	*C. saccharolyticus* DSM8903	75	5	8.9	0.011	2.43	28.2	(Ozdemir *et al.*, 2012)

***Thermotoga maritima* endoglucanases**

As shown in **(Figure 30 and APPENDIX II)**, endoglucanases from *T. maritima* DSM3109 were shown to have the most consistently high specific activities on CMC. The multitude of endoglucanases secreted from this marine organism may serve as empirical evidence for the heterogeneity of endoglucanases, potentially in regards to their penetrative properties and bond cleavage specificity. Moreover, the organism does not secrete any crystalline cellulose degrading enzymes, which paints a picture of the amorphous nature of lignocellulosic substrates it has evolved to depolymerise.

T. maritima MSB8 encodes two GH12 endoglucanases, celA (*Tm*cel12A) and celB (*Tm*celB), in a co-transcribed genetic organization that involves the ATG start codon of *Tm*celB overlapping with the last glutamate codon and stop codon of *Tm*celA. Although their nucleotide sequences are 58% identical, with even similar primary amino acid structures, only *Tm*celB has an amino-terminal signal peptide (Bronnenmeier *et al.*, 1995). Moreover, although *Tm*celA and *Tm*celB presented similar temperature optima and substrate specificity, they differed significantly in their pH-activity profiles and long-term thermostability (Bronnenmeier *et al.*, 1995). Therefore, the strategy to test the heterologous secretion of both GH12 enzymes may be useful in discerning differences in their compatibility in a *G. thermoglucosidasius* TM242 CBP system, especially given their formidable specific activities on amorphous CMC **(Table 12)**.

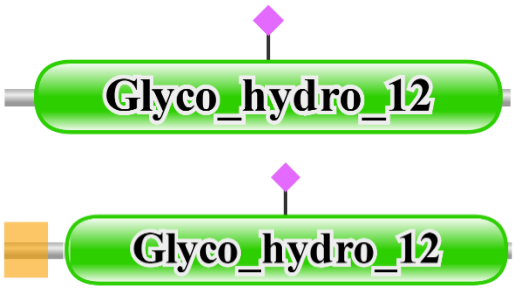


Figure 31 - Annotated protein domain structure of *Thermotoga maritima endoglucanases* A and B, both composed of glycosyl hydrolase family 12 catalytic domains, with *Tm*celB containing an N-terminal signal peptide (yellow box). Active site residues are marked as purple diamonds.

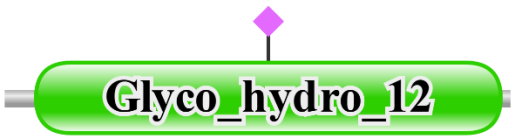


Figure 32 - Annotated protein domain structure of *Thermotoga maritima endoglucanase* 5A, composed exclusively of a glycosyl hydrolase family 5 catalytic domain. Active site residues are marked as purple diamonds.

T. maritima MSB8 also encodes the GH5 endoglucanase *Tm*cel5A, which exhibits dual-activity against both glucan- and mannan-based polysaccharides, which is uncommon for GH5 enzymes (Pereira *et al.*, 2010). Intriguingly, structural studies have suggested that the *Tm*cel5A active site with a wider groove can accommodate not only linear oligosaccharides but also branched ones (Wu *et al.*, 2011).

For expression in *G. thermoglucosidasius* TM242, primer pairs 36(F)/37(R), 38(F)/39(R) and 40(F)/41(R) were designed to facilitate the amplification of *Tm*celA, *Tm*celB and *Tm*cel5A MCS fragments from *T. maritima* DSM3109 genomic DNA, respectively, and subsequently cloned into the expression vector pUCG4.8 downstream of the strong constitutive RPLS promoter and xylanase signal peptide.

Clostridium thermocellum Cel9I

One noteable candidate from the minority of non-cellulosomal cellulases from *C. thermocellum* DSM1237 is endoglucanase *Ct*celI, a GH9 family enzyme, which has been characterised as a processive endoglucanase. In addition to its high specific activity on amorphous CMC (129.2 U/mg), *Ct*celI exhibits a weak ability to degrade crystalline cellulose (Gilad *et al.*, 2003). Moreover, the study identified a 2.1-fold degree of synergism between the processive endoglucanase CelI and the aforementioned non-reducing end exoglucanase CelY. Therefore, the addition of a processive endoglucanase to the envisioned *G. thermoglucosidasius* CBP system can potentially increase its efficacy dramatically.

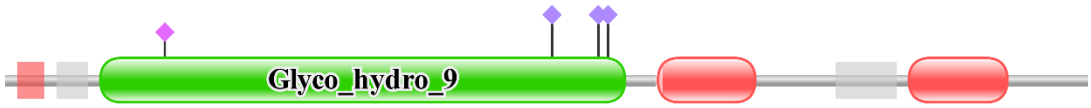


Figure 33 - Annotated protein domain structure of *Clostridium thermocellum* processive endoglucanase I, consisting of an N-terminal signal peptide (red box), glycosyl hydrolase family 9 catalytic domain (green oval) and dual dockerin domains (red ovals). Active site residues are marked as purple diamonds and unstructured regions are marked by grey boxes.

For optimal expression in *G. thermoglucosidasius* TM242, the nucleotide sequence of *Ct*cel9I was codon-harmonised using the *G. thermoglucosidasius* TM242 genome, and synthesised in three parts (*GeneArt DNAstrings/Life Technologies)*. The MCS fragments were subsequently cloned into the expression vector pUCG4.8, downstream of the strong constitutive RPLS promoter and xylanase signal peptide. Primers 42/43 were designed to facilitate the amplification of the MCS fragment if required.

Caldicellulosiruptor saccharolyticus Cel5SLH

One of the goals of this project was the investigation of S-layer attached cellulases as an alternative to natively supernatant-located cellulases. However, the heterologous expression of cellulolytic macro-protein complexes, such as cellulosomes, is currently

beyond the scope of presently-available genetic engineering capabilities. Nevertheless, cellulases have been characterised that facilitate binding to their native cell surface.

C. saccharolyticus has 11 S-layer homology (SLH) domain proteins, including *Cs*cel5SLH (Csac_0678), a GH5 family enzyme that has been characterised to be a bi-functional endoglucanase/xylanase that binds to the S-layer (Ozdemir *et al.*, 2012). Moreover, the domain structure includes a family 28 CBM, which has been reported to form a cleft shape that accommodates cellooligosaccharides (Tsukimoto *et al.*, 2010). The demonstrated high thermostability of *Cs*cel5SLH is particularly impressive, with 50% of its original activity remaining after 48 h of incubation at 75°C.

Although *Cs*cel5SLH has been reported to have low measureable activity on CMC (8.94 U/mg), especially in comparison to *Ct*cel9I and the *T. maritima* endoglucanases (129-890 U/mg), activity on PASC was markedly greater for *Cs*cel5SLH (2.43 U/mg) compared to *Ct*cel9I (1.00 U/mg). However, it should be emphasised that, unlike commercially available CMC, PASC is prepared by each individual laboratory by the phosphoric acid treatment of crystalline cellulose. Therefore, there are likely to be significant variations in the residual crystallinity and recalcitrance of PASC used by these characterisation studies.

Figure 34 - Annotated protein domain structure of *Caldicellulosiruptor saccharolyticus* surface attached GH5 endoglucanase, consisting of an N-terminal signal peptide (yellow box), glycosyl hydrolase family 9 catalytic domain, carbohydrate binding module family 28 domain (red oval), and tripe SLH domains (blue boxes). Active site residues are marked as purple diamonds, and unstructured regions are marked by grey boxes.

For expression in *G. thermoglucosidasius* TM242, primer pair 44(F)/45(R) was designed to facilitate the amplification of the *Cs*celSLH MCS fragments from *C. saccharolyticus* DSM8903 genomic DNA, and subsequently cloned into the expression vector pUCG4.8 downstream of the strong constitutive RPLS promoter and xylanase signal peptide.

Hemicellulases

Xylan is the major hemicellulose component of *Miscanthus* x *giganteus*. The previous analysis of the *G. thermoglucosidasius* TM242 genome has revealed that the strain encodes the

capabilities for trans-membrane transport and subsequent intracellular degradation of xylooligosaccharides, followed by conventional xylose and alternative pentose sugar metabolism pathways. However, the same analysis revealed the loss of xylan polysaccharide degrading xylanases encoded by other *Geobacillus* strains. Although the insertion of the missing 'K' section of the hemicellulose utilisation (HUS) locus will theoretically introduce the capability for growth of *G. thermoglucosidasius* TM242 on xylan as a sole carbon source, it is likely to result from moderate native gene expression of the xylanase *xynA1* gene, and catabolite repression of gene expression is a potential issue.

Figure 35 - Annotated protein domain structure of *Geobacillus thermoglucosidasius* C56-YS93 xylanase A, consisting of an N-terminal signal peptide (yellow box) and glycosyl hydrolase family 10 catalytic domain. Active site residues are marked as purple diamonds.

Therefore, in order to engineer faster cellular growth on xylan, the heterologous expression of the *Geobacillus* spp. xylanase *xynA1* from a strong and constitutive promoter is required.

Lignocellulosic biomass substrates often contain a small proportion of arabinan**,** and the genomic analysis of the *G. thermoglucosidasius* TM242 HUS locus has revealed encoded capabilities for intracellular arabinofuranose degradation and subsequent transport and metabolism of L-arabinose. However, although heterologous expression and secretion of extracellular arabinases and arabinofuranosidases will likely increase the rate of arabinan utilisation, the focus of this study will remain the utilisation of the major polysaccharide components of *Miscanthus* x *giganteus*: cellulose and xylan.

4.3 Prospection of non-catalytic 'cellulase booster' proteins

Lytic Polysaccharide Monooxygenases

The main class of cellulase boosters able to strongly enhance the activity of cellulases are the Lytic Polysaccharide Monooxygenases (LPMOs), capable of cleaving polysaccharide chains in their crystalline regions through an oxidative mechanism (Forsberg *et al.*, 2011). This mechanism involves the oxidation of C1 carbons of glucose, resulting in breaks in the crystalline cellulose chain and formation of aldonic acid by-product. As expected, LPMOs are

found exclusively in aerobic cellulolytic microorganisms, due to the inherent molecular oxygen requirement for enzymatic activity.

Two LPMOs are encoded in *Thermobifida fusca XY* genome. LPMO10A (Tfu_1268; E7; YP_289329), a 22.7 kDa single-domain AA10, has been characterised as a C1/C4-oxidizer (Arfi *et al.*, 2014). LPMO10B (Tfu_1665; E8), a 44.7-kDa three-domain protein consisting of a AA10, fibronectin 3 and Family 2 CBM domain, has been characterised as a C1-specific oxidiser (Arfi *et al.*, 2014). When assayed for their ability to cleave Avicel in the presence of ascorbic acid as an electron donor, both LPMO10A and LPMO10B released a mixture of soluble sugars. High performance anion exchange chromatography (HPAEC) identified that both enzymes released mixtures with similar compositions, comprising reduced and oxidized cellooligosaccharides (Arfi *et al.*, 2014).

Figure 36 – Annotated protein domain structure of *Thermobifida fusca* XY lytic polysaccharide monooxygenase A, consisting of an N-terminal signal peptide (yellow box) and chitin binding 3 (AA10) domain. Metal binding residues are marked by blue circles.

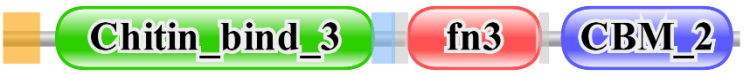


Figure 37 – Annotated protein domain structure of *Thermobifida fusca* XY lytic polysaccharide monooxygenase B, consisting of an N-terminal signal peptide (yellow box) and chitin binding 3 (AA10) domain, fibronectin 3 and family 2 carbohydrate binding module domain. Unstructured regions are marked by grey boxes.

For optimal expression in *G. thermoglucosidasius* TM242, the nucleotide sequences of LPMO10A and LPMO10B were codon-harmonised using the *G. thermoglucosidasius* TM242 genome, and synthesised (*GeneArt/Life Technologies)*. The MCS fragments were subsequently cloned into the expression vector pUCG4.8, downstream of the strong constitutive RPLS promoter. Primer pairs 46(F)/47(R) and 48(F)/49(R) were designed to facilitate the amplification of the LPMO10A and LPMO10B MCS fragments, respectively.

Expansins

Another class of proteins with lignocellulolytic-loosening activity are the aptly named Loosenin family of fungal proteins, with the best characterised loosenin isolated from *Bjerkandera adusta* (Quiroz-Castañeda *et al.*, 2011). However, the protein exhibited low thermostability (< 40°C) and homology analysis identified no homologs in the bacterial domain (Quiroz-Castañeda *et al.*, 2011).

Another class of proteins with lignocellulosic-loosening proteins are the fungal Swollenin proteins from the primary cellulolytic fungus *Trichoderma reesei* and *Aspergillus fumigatus* (Saloheimo *et al.*, 2002, Chen *et al.*, 2010). Homology analysis reveals strong homology to the bacterial expansin-like proteins, but from mesophilic isolates. Although *T. reesei* swollenin does exhibit thermoactivity at 50°C (pH 5.0), the cross-domain differences in post-translational protein modifications may be problematic (Saloheimo *et al.*, 2002).

Annually, there is increasing international interest in the study of *expansins* and *expansin-like* proteins, which serve to de-crystallise the macromolecular structure of lignocellulose. This synergistically facilitates the penetration of catalytic glycosyl hydrolases into the lignocellulosic mileau, and has been shown to significantly increase the hydrolytic activity of cellulolytic mixtures (Sampedro and Cosgrove, 2005).

Figure 38 – Annotated protein domain structure of putative *Clostridium clariflavum* DSM19732 expansin-like protein, consisting of an N-terminal signal peptide (yellow box), dual type I dockerin domains, followed by an unannotated putative expansin-like protein region. Unstructured regions are marked by grey boxes.

However, there is, as of yet, no bacterial homolog of the predominantly fungal proteins, and no thermostable equivalent for characterised bacterial expansin-like proteins. The amino acid sequence of the best characterised expansin-like protein to date, *Bacillus subtilis* EXLX1 (sometimes referred to as YoaJ), was used to perform bi-lateral homology analysis against the NCBI nucleotide collection and NCBI whole genome shotgun contigs (WGS) database. This analysis revealed the presence of a homolog, with 61% sequence identity over 90% sequence coverage, present in the genome of the cellulolytic thermophile *Clostridium clariflavum*

DSM19732. This homolog contains N-terminal type I dockerin domains, and is therefore likely to be cellulosomal. However, a similar homolog is present in the WGS sequence of *Clostridium clariflavum* strain 4-2a, with 62% sequence identity to *B. subtilis* YoaJ, but without the presence of an N-terminal dockerin domains.

LipoP analysis of the translated sequence of this expansin-like protein (*Clocl_1862*), henceforth referred to as CcYoaJ, revealed the presence of an SPII signal peptide between 20/21 aa. Primer pair 50(F)/51(R) were designed to amplify the MCS fragment for cloning into an expression vector.

Serpins

There is increasing evidence for the putative proteolytic activity of *Geobacillus* culture supernatants (Chen *et al.*, 2004, Güracar, 2011, Hawumba *et al.*, 2002, Iqbal *et al.*, 2015, Miyake *et al.*, 2005, Zhu *et al.*, 2007). In fact, the *G. thermoglucosidasius* C56-YS93 and TM242 genomes both encode for eight putatively-secreted putative proteases with predicted Pfam peptidase domains **(Table 13)**.

This may serve to be problematic in strains engineered to secrete heterologous glycosyl hydrolases, which would be susceptible to proteolytic degradation. The long-term productively-efficient strategy to circumvent this potential issue is the systematic knockout of the putative protease genes, which has been performed in the pharmaceutical protein producing *B. subtilis* WB800 strain (Westers *et al.*, 2004).

Table 13 - **Predicted peptidases in translated *Geobacillus thermoglucosidasius* genomes.** The GeoSec score is a consolidated scoring system for secretion protein prediction (Chapter 6). A threshold of 0.6 was chosen to isolate putative extracellular peptidases. Pfam domains were annotated using the batch Pfam annotation tool (*Sanger*)

Protein Name	11955 Identifier	C56-YS93 Identifier	Y41.MC1 Identifier	Pfam Domains	GeoSec Score
Subtilisin	**peg3497**	YP_004586418.1	YP_003987734.1	**Peptidase**_S8; PPC	0.88516
β-lactamase	**peg1167**	YP_004588444.1	YP_003989718.1	Beta-lactamase; **Peptidase**_M6	0.8386

Oligoendopeptidase F	**peg1500**	YP_0045 88205.1	YP_0039 89457.1	**Peptidase**_M3_N; **Peptidase**_M3	0.80656
Peptidoglycan lytic protein P45	**peg3363**			**Peptidase**_M23	0.78984
Serine-type D-Ala-D-Ala carboxypeptidase	**peg2510**	YP_0045 87426.1	YP_0039 88664.1	**Peptidase**_S11; PBP5_C	0.76714
Glutamyl endopeptidase	**peg0108**				0.74956
Carboxyl-terminal protease	**peg3360**	YP_0045 86523.1	YP_0039 87830.1	PDZ_2; **Peptidase**_S41; PG_binding_1	0.71888
Cell wall endopeptidase, family M23/M37	**peg3208**			**Peptidase**_M23	0.69332
Hypothetical protein Geoth_0962		YP_0045 87041.1	YP_0039 88317.1	Colicin; Hemerythrin; Spore_III_AB; **Peptidase**_M75; NVEALA	0.69332
Serine-type D-Ala-D-Ala carboxypeptidase		YP_0045 86172.1	YP_0039 87477.1	**Peptidase**_S11; LTD; PBP5_C	
peptidase M23		YP_0045 86520.1	YP_0039 87827.1	Poty_PP; IncA; AAA_13; CALCOCO1; TPR_MLP1_2; APG6; ATG16; Sec8_exocyst; **Peptidase**_M23; TBPIP; Seryl_tRNA_N; Reo_sigmaC; GAS; CENP-F_leu_zip; YlqD; TMF_DNA_bd; Seryl_tRNA_N; DASH_Dam1	

However, for the purpose of this study, the expression and secretion of the serpin class of serine-protease inhibitor proteins may provide a short-run solution. These proteins serve as broad inhibitors of serine proteases, functioning as metastable suicide substrates for their cognate proteases, and are thought to play a key role in protecting the cellulosome of some Clostridia species from protease attack (Schwarz and Zverlov, 2006). Rather than folding to their minimum free energy structures, serpins fold to a metastable state (Tsutsui *et al.*, 2006). Proteolytic cleavage of its flexible reactive centre loop brings about a massive conformational change that results in the insertion of the flexible reactive centre loop into the central beta-sheet of the serine-protease **(Figure 39)**. Thus, the inactivation of the target protease is permanent via a serpin-protease covalent-bond interaction.

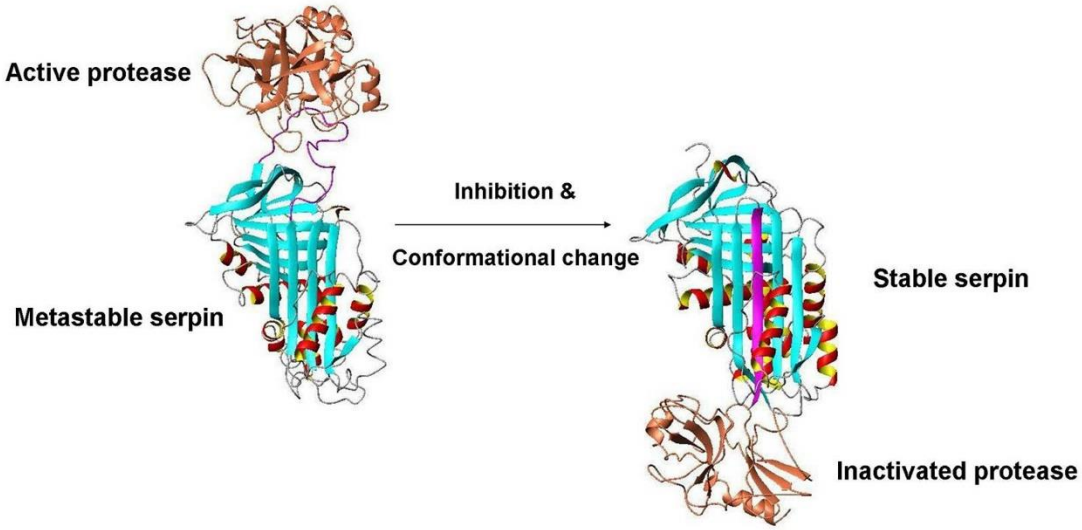


Figure 39: **Mechanism of permanent serine protease inhibition by metastable serpin protein** (*adapted from University of Maryland: http://faculty.rx.umaryland.edu/pwintrode/research-projects/* [Accessed on 21st March 2016])

A characterised serpin of interest, the cellulosomal PinA protein expressed by *C. thermocellum* DSM1237, was chosen. Homology analysis using the *C. thermocellum* DSM1237 PinA protein sequence (ABN51432.1) revealed no serpin homologs in thermophilic *Bacilliceae*. This is to be expected since Bacilli, Geobacilli, and other genera secrete proteases and the secretion of serine-protease inhibitors would be functionally counter-intuitive. A similar homology analysis identified 94 serpin homologs from *Clostridiaceae*, including a non-cellulosomal serpin paralog (ABN52502.1) in the *C. thermocellum* DSM1237 genome. LipoP analysis of the translated sequence of this non-cellulosomal serpin (*Cthe_1270*), henceforth referred to as *Ct*PinB, revealed the presence of an SPII signal peptide between 20/21 aa.

Primers 52(F)/53(R) were designed to amplify the MCS fragment for cloning into the pUCG_Prom3_SigPep1 expression vector.

4.5 Development of a synthetic-biology *Geobacillus spp.* expression platform

The construction of the three-part *E. coli-Geobacillus* shuttle vector pUCG3.8 has already facilitated the expression and secretion of endoglucanases (Bartosiak-Jentys *et al.*, 2013). However, for a combinatorial approach to the expression of several glycosyl hydrolases, there is a need for further improvements in the clonal efficiency in *E. coli*, both at construct assembly and the transformation process.

Improvement of clonal frequency in *E. coli* (pUCG4.8)

The efficiency of a plasmid's selection marker in the cloning organism is one of the primary determinants of transformation efficiency. Although, the thermostable Kanamycin resistance marker (Liao and Kanikula, 1990) has shown functionality in *E. coli* BioBlue cells, the introduction of a robust *E. coli* resistance marker may increase the efficiency of the transformation process. To this end, an *E. coli*-compatible 1,007 bp Ampicillin-resistance marker was amplified from the holding vector pJET1.2 (*Fermentas*), using the primer pair 60(F)/61(R), and introduced between the *Zra*I and *Bss*HII sites of the pUCG3.8 vector backbone **(Figure 40)**, resulting in the 4,751 bp vector pUCG4.8.

To test for an improvement in cloning efficiency, *E. coli* DH5α and Bioblue cells were transformed with a ligation of a GFP-expressing cassette into either pUCG3.8 or pUCG4.8, followed by an overnight incubation on kanamycin-supplemented and ampicillin-supplemented rich agar. All resultant colonies observed were fluorescent. However, as shown in **Table 14**, there was more than a 41-fold and 4-fold increase in positive *E. coli* DH5-α and Bioblue colonies, respectively, when a pUCG4.8 transformation was selected on ampicillin compared to a pUCG3.8 transformation selected on kanamycin

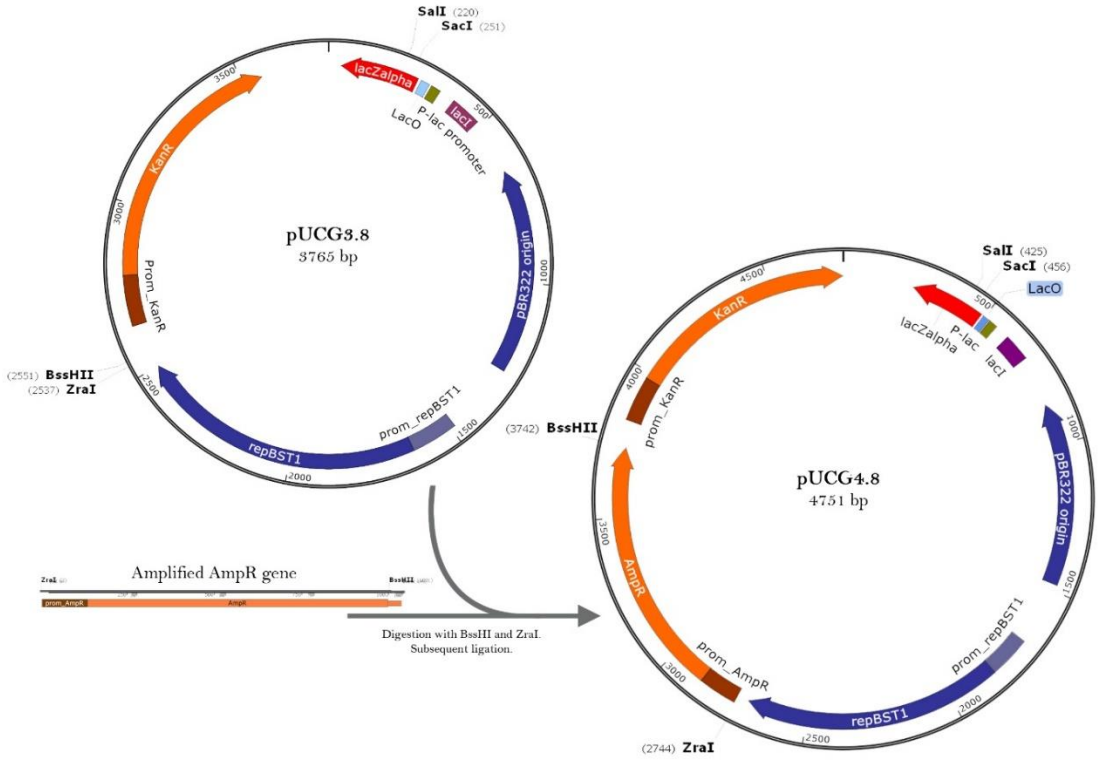


Figure 40 - Construction of ampicillin resistance expressing shuttle vector pUCG4.8, from pUCG3.8.

Table 14 – Fluorescent colonies resultant from transformation of GFP-expression cassette ligated into pUCG3.8 and pUCG4.8.

Ligation	Cells	Fluorescent colonies on LBA100 plate	Fluorescent colonies on LBK30 plate
pUCG3.8_RPLSGFP	*E. coli* DH5-α	-	3 ± 1
pUCG4.8_RPLSGFP	*E. coli* DH5-α	125 ± 12	2 ± 1
pUCG3.8_RPLSGFP	*E. coli* BioBlue	-	36 ± 15
pUCG4.8_RPLSGFP	*E. coli* BioBlue	143 ± 24	28 ± 9

As a result, pUCG4.8 was used as the plasmid backbone for the heterologous expression of glycosyl hydrolases in *G. thermoglucosidasius.*

A simplified and robust clonal method for the introduction of MCS fragments into pUCG4.8

For comparisons of glycosyl hydrolases from each enzyme class, individual MCS fragments must be introduced downstream of the same constitutive promoter and native signal peptide pair. To date, this has been done by conventional restriction digest and ligation of the amplified MCS fragment into a pUCG4.8_PromRPLS_SigPep1_MCS vector, whereby PromRPLS denotes the constitutive RPLS promoter, SigPep1 denotes the *G. thermoglucosidasius* C56-YS92 xylanase signal peptide, and MCS denotes any mature protein coding sequence fragment.

However, similar to the sfGFP-based cloning discussed in **Section 3.1,** to circumvent the issues that complicate the conventional construction of expression plasmids, the GoldenGate method of cloning was exploited to construct a variant of pUCG4.8 that facilitates seamless and targeted insertion of MCS fragments downstream of the constitutive PromRPLS and SigPep1. To target an MCS fragment downstream of the signal peptide, the forward GoldenGate end consists of a *Bsa*I recognition site followed by an *Ngo*MIV site, which facilitates its insertion at the *Ngo*MIV site immediately downstream of a Signal Peptide-coding fragment. A universal stop codon '`TAA`' is used for all MCS fragments, and is present in the reverse GoldenGate end, which consists of a *Bsa*I recognition site followed by the 6-bp sequence '`GCGTTA`' on the reverse strand. The forward and reverse GoldenGate ends are added to the 5' ends of the forward and reverse primers used for the amplification of the MCS fragment, respectively **(Figure 41)**.

To circumvent any potential mutational variance between the several expression constructs, a version of pUCG4.8 containing the constitutive RPLS promoter, *Geobacillus*-compatible signal peptide 1 and a GoldenGate-ready section was constructed. Once digested with *Bsa*I, this vector generates the precise compatible ends for unidirectional ligation of the MCS fragments containing the corresponding GoldenGate ends downstream of SigPep1. To facilitate this, a silent site-directed mutation of a *Bsa*I site present in the ampicillin resistance marker of pUCG4.8, was required. Using the primer pairs 4(F)/62(R) and 63(F)/3(R), the vector pUCG4.8_PromRPLS_sfGFP was amplified in two parts (1,781-bp and 3,179-bp) and assembled by a GoldenGate reaction into pUCG4.8_PromRPLS_SigPep1_GoldenGate.

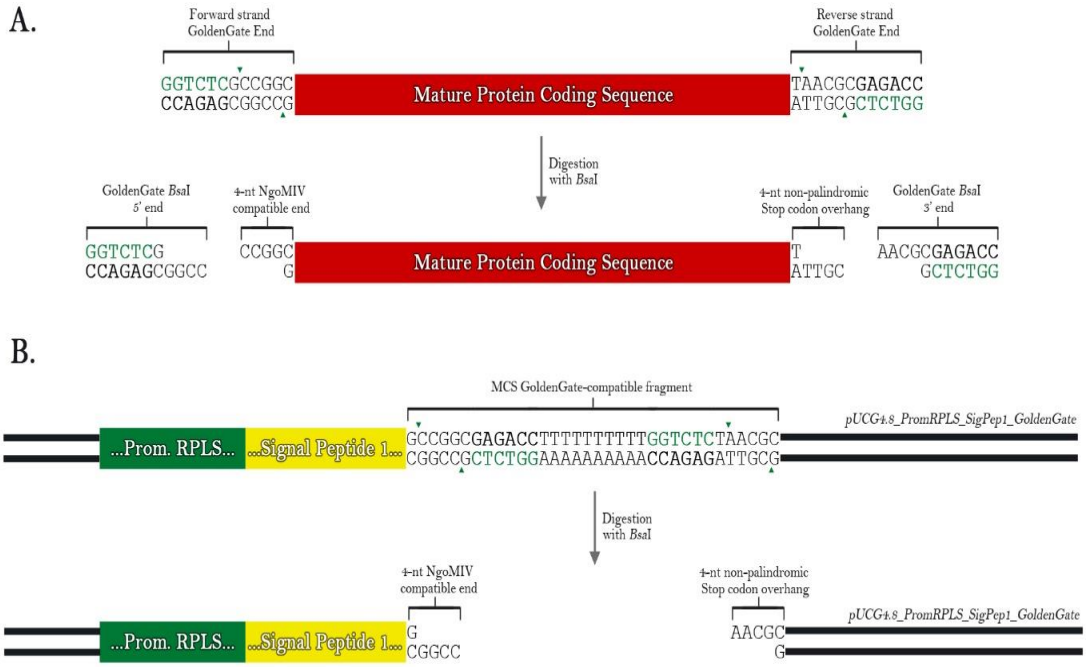


Figure 41: GoldenGate-based insertion of mature protein coding sequence (MCS) fragments into the shuttle expression vector pUCG4.8_PromRPLS_SigPep1_GoldenGate. (A) The forward and reverse GoldenGate ends that flank the MCS fragment are designed such that *Bsa*I digestion at the 5' *Bsa*I site (green lettering) results in the generation of specific 4-nt overhangs. (B) *Bsa*I digestion of the MCS GoldenGate-compatible fragment, which has been cloned downstream of a xylanase signal peptide (Signal Peptide 1) and strong and constitutive RPLS promoter, facilitates the generation of complimentary 4-nt overhangs to those generated on the MCS fragment. Ligation of these overhangs is simultaneously mediated by T4 DNA ligase.

In conclusion, the presented glycosyl hydrolases provided a starting panel of prospective cellulases for the desired *G. thermoglucosidasius* CBP organism. Although endoglucanases are the most widely studied type of cellulose, the importance of introduced exoglucnase-seceting *G. thermoglucosidasius* is paramount, especially given that no member of the *Geobacillus* genus has been isolated expressing an effective exoglucanase. Nevertheless, this panel of glycosyl hydrolases was sufficient to facilitate proof-of-concept expression studies outlined in Chapter 5.

Chapter 5: Characterisation of cellulase-secreting *G. thermoglucosidasius* TM242

Effective CBP engineering in *G. thermoglucosidasius* TM242 involves iterative optimisation of engineered cellulolytic capabilities. Prior to this project, there was no convenient high-throughput screening process for effective cellulases and no effective process for following microbial growth on insoluble substrates. The metabolic burden of cellulase expression and secretion is an issue that needs evaluating, but there was no available data on heterologous protein secretion in *G. thermoglucosidasius* TM242. Moreover, the putative synergistic activities of the *C. clariflavum* YoaJ expansin (*Cc*YoaJ) had not been investigated. Therefore, it was the remit of phase II of this project to address these concerns, issues and knowledge gaps as a foundation for future optimisation of a *G. thermoglucosidasius* TM242 CBP process.

5.1 Alteration of the DNS protocol for higher-throughput screening of cellulase activity.

The combinatorial nature of CBP engineering, with multiple variations of promoters, signal peptides and mature protein coding sequence (MCS) fragments, required high-throughput testing and comparison of cellulase activities in a number of strains. The conventional and most widely accepted method for testing cellulolytic activity is the dinitrosalicylic acid (DNS) assay method, a discontinuous assay that measures the concentration of reducing sugars and other reducing molecules released in a set time, which reduce DNS to form 3-amino-5-nitrosalicylic acid (ANS). Since the hydrolysis of glycosidic bonds of polysaccharides results in the release of a reducing-end, the activity of glycosyl hydrolases can be measured by an increase in ANS, which absorbs light strongly at 540 nm.

The discontinuity of the DNS assay causes complications for comparisons of cellulolytic *Geobacillus* cultures, since the process from supernatant sampling to measurement data is time-consuming. This process involves the transfer of supernatant and solutions from 50 ml centrifuge tubes to 15 ml centrifuge tubes, and finally to 1.5 ml centrifuge tubes, using a single-channel pipette **(Conventional DNS Assay Protocol)**.

Conventional DNS Assay Protocol
Sample 3-5 ml of culture into 50 ml centrifuge tube; centrifuge at 4,000 rpm for 1 hour
Carefully transfer 1 ml culture supernatant into 15 ml centrifuge tube; add 1 ml 2.5% CMC
Incubate at 60°C in heat block for chosen reaction time (default: 1 hour)
Transfer 200 µl into 1.5 ml centrifuge tube containing 400 µl 3:1 DNS solution
Incubate at 100°C in heat block for 20 minutes; test absorbance at 540 nm

This is not only time-consuming for the researcher, but can skew the results of the experiment if too many samples are being processed simultaneously, as the reaction time for the first sample may be significantly longer than that of the last reaction. Therefore, for higher-throughput analysis, amendment of the DNS assay was necessary to allow simultaneous testing of numerous samples, and the use of multi-channel pipettes to reduce time differentials between samples.

Amendments to the conventional DNS assay were tested using Novozyme's Viscozyme L enzyme mixture on carboxymethylcellulose (CMC) **(DNS_Prot2 and DNS_Prot3 below)**. DNS_Prot2 had only minor amendments to the conventional DNS_Prot1; the volume of the hydrolysis reaction was reduced to facilitate the universal use of 1.5 ml centrifuge tubes. In contrast, DNS_Prot3 had multiple variations from DNS_Prot3; the volume of both the polysaccharide hydrolysis reaction and the DNS-reduction reaction was reduced to facilitate the use of PCR tubes and thermocyclers, and the use of multichannel pipettes for rapid and reliable transfer of solutions.

There was no significant change in measured specific activities between the conventional and amended DNS protocols for Viscozyme L on CMC **(Figure 42)**; both 0.001% and 0.0005% dilutions of Viscozyme L gave an average of 1053 ± 21.5 µmol glucose equiv./min/ml.

DNS_Prot2
Sample 1-1.3 ml of culture into 1.5 ml centrifuge tube; centrifuge at 6,000g for 1 hour
Carefully transfer 500 µl culture supernatant into 1.5 ml centrifuge tube; add 500 µl 2.5% CMC
Incubate at 60°C in heat block for chosen reaction time (default: 1 hour)
Transfer 200 µl into 1.5 ml centrifuge tube containing 400 µl 3:1 DNS solution
Incubate at 99°C in heat block for 20 minutes; test absorbance at 540 nm

DNS_Prot3
Sample 1-1.5 ml of culture into 1.5 ml centrifuge tube; centrifuge at 6,000g for 1 hour
Carefully transfer 50 µl culture supernatant into 250 µl PCR tube; add 50 µl 2.5% CMC
Incubate at 60°C in thermocycler for chosen reaction time (default: 1 hour)
Transfer 60 µl into a 250 µl PCR tube containing 120 µl DNS solution
Incubate at 99°C in thermocycler for 30 minutes; test absorbance at 540 nm

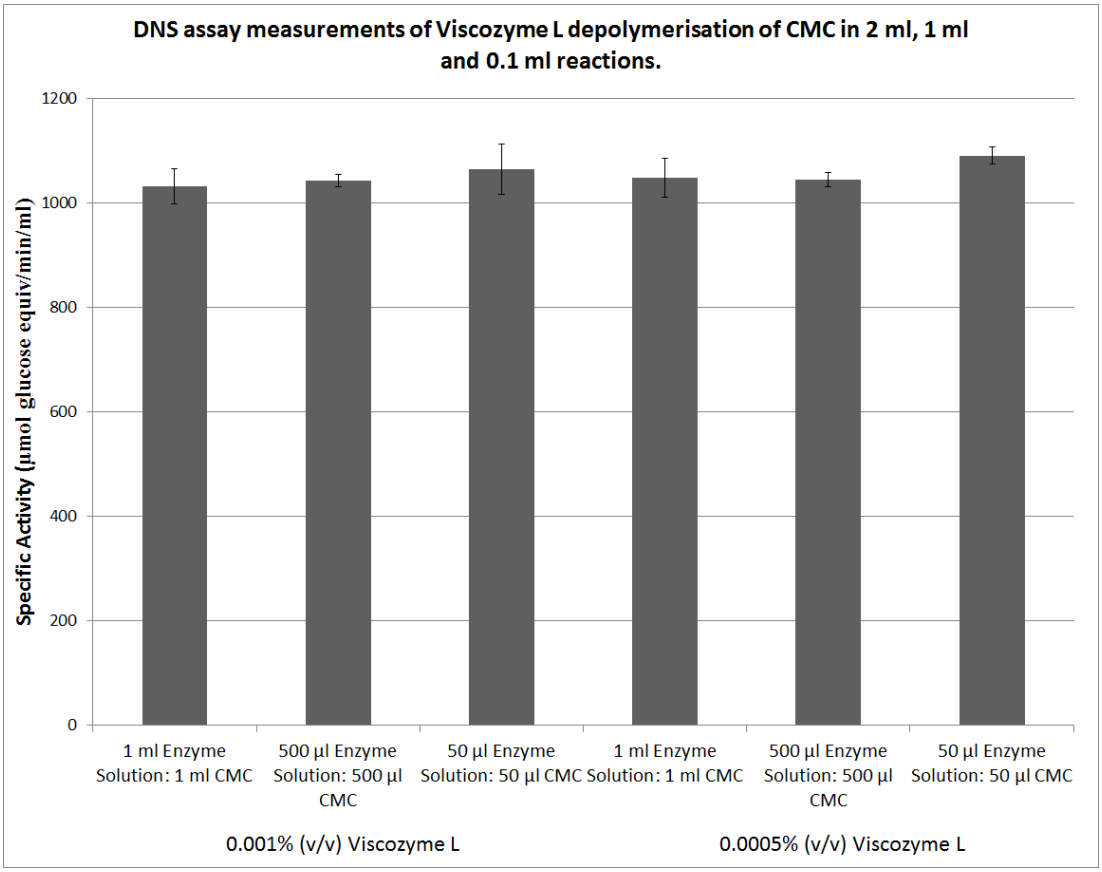


Figure 42 – Measurement of Viscozyme L degradation of CMC using the conventional and amended DNS assay protocols. Reaction times were kept constant at 15 minutes. 0.001% and 0.0005% (v/v) Viscozyme L were used in 1 ml, 500 µl, and 50 µl volumes for subsequent hydrolysis of 2.5% CMC, and dilutions were subsequently factored into measurements of glucose equivalent concentrations.

The amendment of this protocol paved the way for extended *in vitro* analysis of glycosyl hydrolase activities, and the design of combinatorial cellulase assays that were not previously possible.

5.2 Specific activities of cellulase-secreting *Geobacillus thermoglucosidasius* TM242 strains.

Since the primary goal of this project was the secretion of active heterologous cellulases, the specific cellulase productivity during bacterial growth and in stationary phase was of particular importance. To investigate this, individual endoglucanase-secreting and exoglucanase-secreting *G. thermoglucosidasius* TM242 strains (described in **section 4.2**) were grown in ASM minimal media containing 0.05% Yeast Extract and 2% Glycerol as sole carbon sources to minimise the presence of assay-interfering reducing-sugars in culture

supernatants. Specific activities for the degradation of CMC or phosphoric-acid swollen cellulose (PASC) were measured in culture supernatants from endoglucanase- and exoglucanase-secreting strains, respectively, initially at a culture OD_{600nm} of 1 and at 500 minutes post-inoculation.

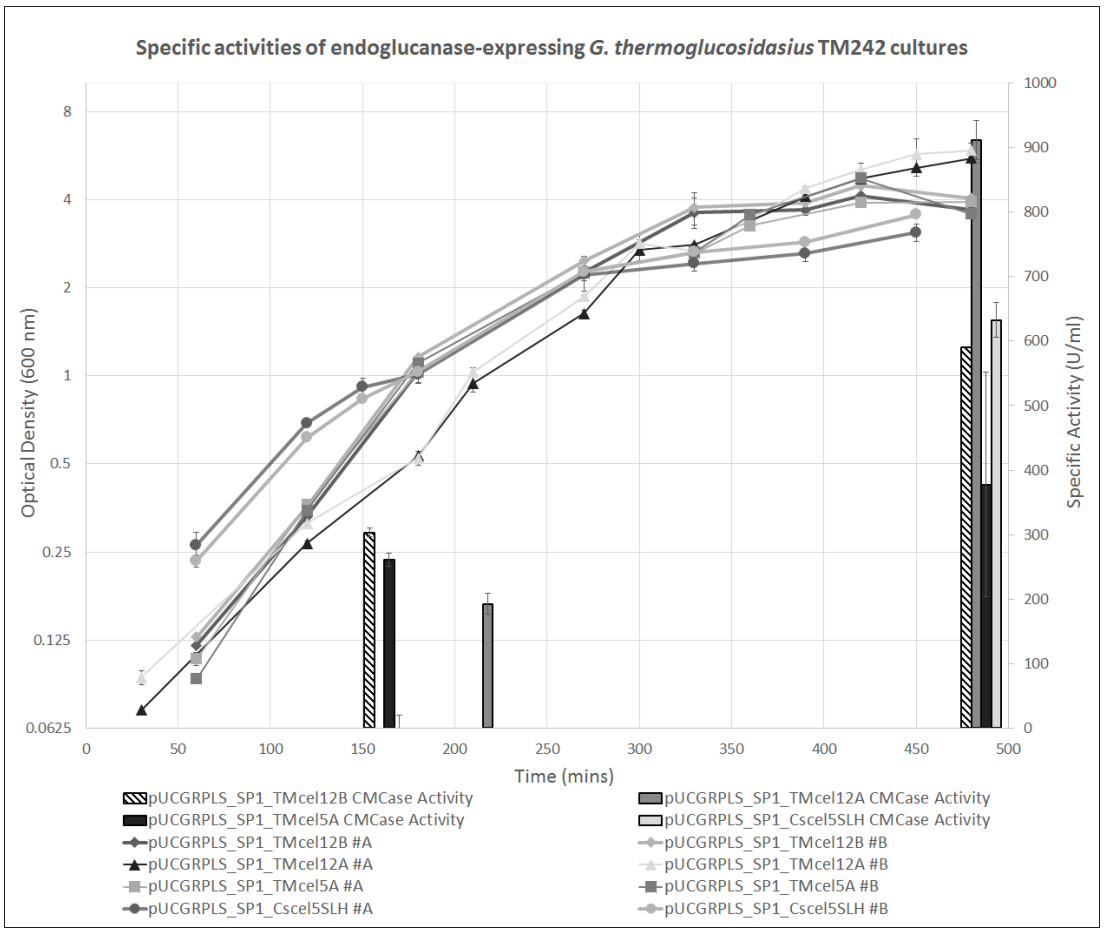


Figure 43 – CMC-degrading specific activities of endoglucanase-secreting *G. thermoglucosidasius* TM242 strains. Supernatants were sampled from duplicate aerobic cultures of *Tm*cel12A-, *Tm*cel12B-, *Tm*cel5A-, and *Cs*cel5SLH-secreting *G. thermoglucosidasius* TM242 strains grown on minimal media. Supernatants were collected when culture OD_{600nm} reached 1 AUs and at 500 minutes post-inoculum. Supernatants were subsequently tested by DNS Protocol 3 for CMC-degrading activities (1.5 hour reaction time).

All endoglucanase-secreting *G. thermoglucosidasius* TM242 strains reached high culture densities, with no significant metabolic burden as reflected by growth rate and maximum culture OD_{600nm} attained **(Figure 43)**; the culture OD_{600nm} for all four endoglucanase-secreting cultures were around 2.42 - 3.76 after 300 minutes of growth. However, variations in the growth profiles to achieve this maximum culture OD_{600nm} were observed.

The close alignment of *Tm*cel12B- and *Tm*cel5A-secreting *G. thermoglucosidasius* TM242 culture profiles to that of the negative control strain without a mature protein coding sequence (MCS) provided a benchmark for comparison with the culture profiles of other cellulase-secreting strains **(Figure 44)**. Clearly the expression and secretion of *Tm*cel12B and *Tm*cel5A, both natively extracellular enzymes, had limited metabolic burden on their host strain. This was not completely unexpected, given the relatively small open reading frames (ORFs) of these two *T. maritima* MS8 enzymes (*Tm*cel12B = 846 bp, *Tm*cel5A = 1044 bp).

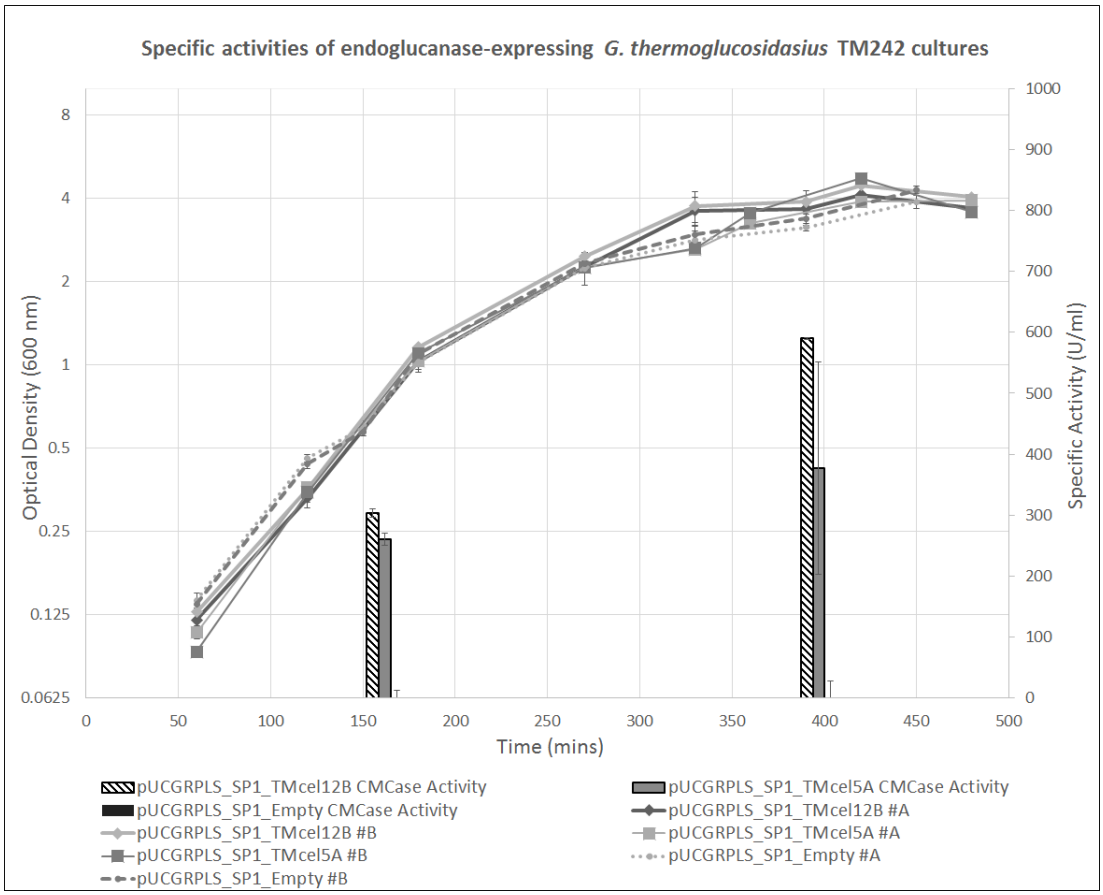


Figure 44 – CMC-degrading specific activities of endoglucanase-secreting *G. thermoglucosidasius* TM242 strains relative to empty vector-expressing *G. thermoglucosidasius* TM242. Supernatants were sampled from duplicate aerobic cultures of *Tm*cel12B-, and *Tm*cel5A-secreting *G. thermoglucosidasius* TM242 strains grown on minimal media, and from a *G. thermoglucosidasius* TM242 strain expressing an empty MCS fragment. Supernatants were collected when culture OD_{600nm} reached 1 AUs and at 500 minutes post-innoculum. Supernatants were subsequently tested by DNS Protocol 3 for CMC-degrading activities (1.5 hour reaction time).

At similar culture OD_{600nm}, culture supernatants from *Tm*cel12B-secreting *G. thermoglucosidasius* had consistently greater CMC-degrading specific activities than its

*Tm*cel5A counterpart, but this might simply reflect the higher specific activity of *Tm*cel12B on CMC (890 U/mg) compared to *Tm*cel5A (616 U/mg).

Although *Tm*cel12A-secreting *G. thermoglucosidasius* cultures exhibited the greatest final culture OD_{600nm} of 5.717 ± 0.17 AU, its growth rate was slower than that of *Tm*cel12B- and *Tm*cel5A-secreting strains. Supernatant collected from *Tm*cel12A-secreting strains at this final culture OD_{600nm} exhibited the highest level of CMC-degrading activity measured (911 U/ml).

A deleterious effect on culture growth was observed for *G. thermoglucosidasius* TM242 secreting *Cs*cel5SLH, the putative cell-surface display endoglucanase, which exhibited a slower growth rate and lower final culture OD_{600nm} than its *T. maritima* endoglucanase-secreting counterparts. The *Cs*cel5SLH construct was actually expressed from the weak pUP2n38 promoter **(Section 5.2)**, rather than the strong pRPLS promoter that drove expression of the other endoglucanases, so expression should have presented less of a burden. This further emphasises the burden of *Cs*cel5SLH secretion, and this may due to potentially deleterious effects caused by endoglucanase attachment to the cell surface of the organism. This effect may be mediated either by direct disruption of S-layer stability by the heterologous hydrolytic function of *Cs*cel5SLH, although this seemed unlikely given its native attachment to the *C. saccharolyticus* S-layer, or due to competition for space on the *G. thermoglucosidasius* S-layer causing structural instabilities.

Intriguingly, no detectable CMC-degrading activities were observed in supernatants from *Cs*cel5SLH-secreting *G. thermoglucosidasius* collected at a culture OD_{600nm} of 1.0, but supernatant collected at a culture OD_{600nm} of 3.3 ± 0.3 exhibited 632.4 U/ml of CMC-degrading specific activity. Although the initially undetectable activity might has been attributed to weak expression from the pUP2n38 promoter, this does not account for the high activity observed after 4.5 hours. The likely explanation, based on the S-layer binding characteristics of the endoglucanase, was that at the lower OD, there was little free *Cs*cel5SLH present in the supernatant. However, *G. thermoglucosidasius*, and other members of the *Geobacillus* genera, exhibit notable autolysis, and this probably contributed to a dramatic increase in *Cs*cel5SLH-displaying cell-wall fragments present in the latter supernatants, leading to the dramatic rise in measured CMC-degrading activity.

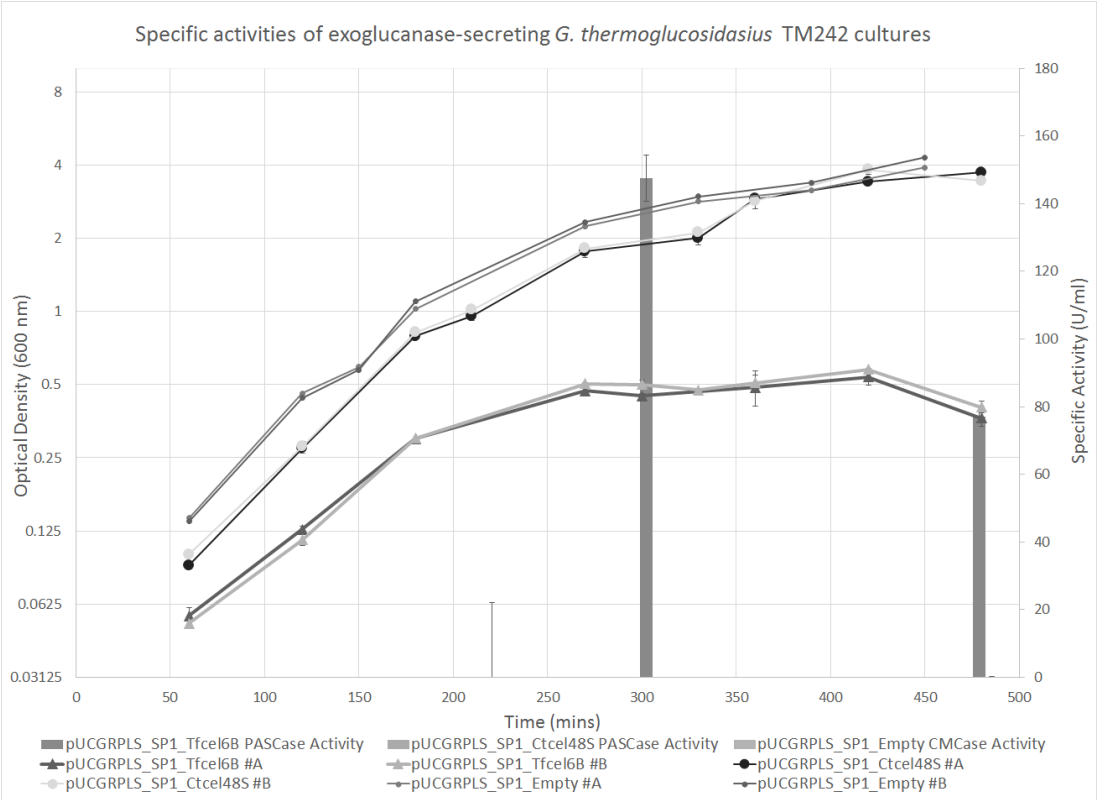


Figure 45 – PASC-degrading specific activities of exoglucanase-secreting *G. thermoglucosidasius* TM242 strains. Supernatants were sampled from duplicate aerobic cultures of *Ct*cel48S- and *Tf*cel6B-secreting *G. thermoglucosidasius* TM242 strains grown on minimal media, and from a *G. thermoglucosidasius* TM242 strain expressing an empty MCS fragment. Supernatants were collected when culture OD_{600nm} reached 1 AUs and at 500 minutes post-innoculum. Supernatants were subsequently tested by DNS Protocol 3 for CMC-degrading activities (1.5 hour reaction time).

Remarkably, *G. thermoglucosidasius* cultures secreting the exoglucanase *Ct*cel48S, which is almost 3 times larger in primary amino acid structure than the endoglucanase *Tm*cel12B and *Tm*cel5A, reached similar culture OD_{600nm} to the endoglucanase-secreting strains, with no significant difference in growth rates. In contrast, *G. thermoglucosidasius* TM242 secreting the exoglucanase *Tf*cel6B exhibited notable stuttering of achievable culture OD_{600nm}; the highest achievable culture OD_{600nm} was at 0.577 AU, before reducing thereafter **(Figure 45)**. Since *Tf*cel6B had been reported to be heavily glycosylated natively, the heterologous glycosylation of the exoglucanase may have presented a great degree of metabolic burden. Since the activity of *Tf*cel6B was likely to be dependent on the glycosylation of the linker peptide connecting its catalytic module and a CBM3 domain, as has been reported for other

cellulases (Zhang *et al.*, 1995, von Ossowski *et al.*, 2005), the *N*-glycosylation of *Tf*cel6B may present a further obstacletowards the conveyance of extracellular cellulolytic activity.

Nevertheless, culture supernatants from *Tf*cel6B-secreting *G. thermoglucosidasius*, sampled as culture OD_{600nm} stabilised at 0.47 ± 0.03 AU, exhibited impressive PASC-degrading specific activities of 147.6 U/ml. However, when culture supernatant was sampled 3 hours later, the measured specific activity on PASC had reduced to 77 U/ml. If this reduction in extracellular PASC-degrading activity was due to the removal of active *Tf*cel6B it may be some indication towards the predicted proteolytic digestion of extracellular proteins in *G. thermoglucosidasius* TM242 **(Section 4.3)**. A counterargument against the hypothesis of proteolytic cleavage is that no reduction in cellulolytic activity was observed with endoglucanase-secreting strains, which secrete the same native extracellular peptidases. However, exoglucanases are typically more flexible than endoglucanases, and flexible regions in proteins are more susceptible to proteolytic cleavage.

5.3 Investigation of synergistic activities between secreted cellulases.

In order to engineer an effective and efficient cellulolytic system, it was important to limit the number of glycosyl hydrolases secreted, thus reducing the metabolic burden of the heterologous expression of these enzymes, and likely improving ethanol yield. One approach in achieving this was the exploitation of any synergism that arose between different glycosyl hydrolases, as had been reported in native cellulolytic systems of *T. fusca* XY and *C. thermocellum* DSM1237.

To investigate any synergism between members of our panel of cellulases, supernatants were isolated at a culture OD_{600nm} of 2 AU from cellulase-secreting *G. thermoglucosidasius* TM242 cultures grown on 2xTY rich media, and tested for their individual and combinatorial activities in the degradation of PASC. In order for synergism to be identified between two cellulase supernatants, the depolymerisation of PASC in combination with each other was required be greater than the additive activities of the individual culture supernatants.

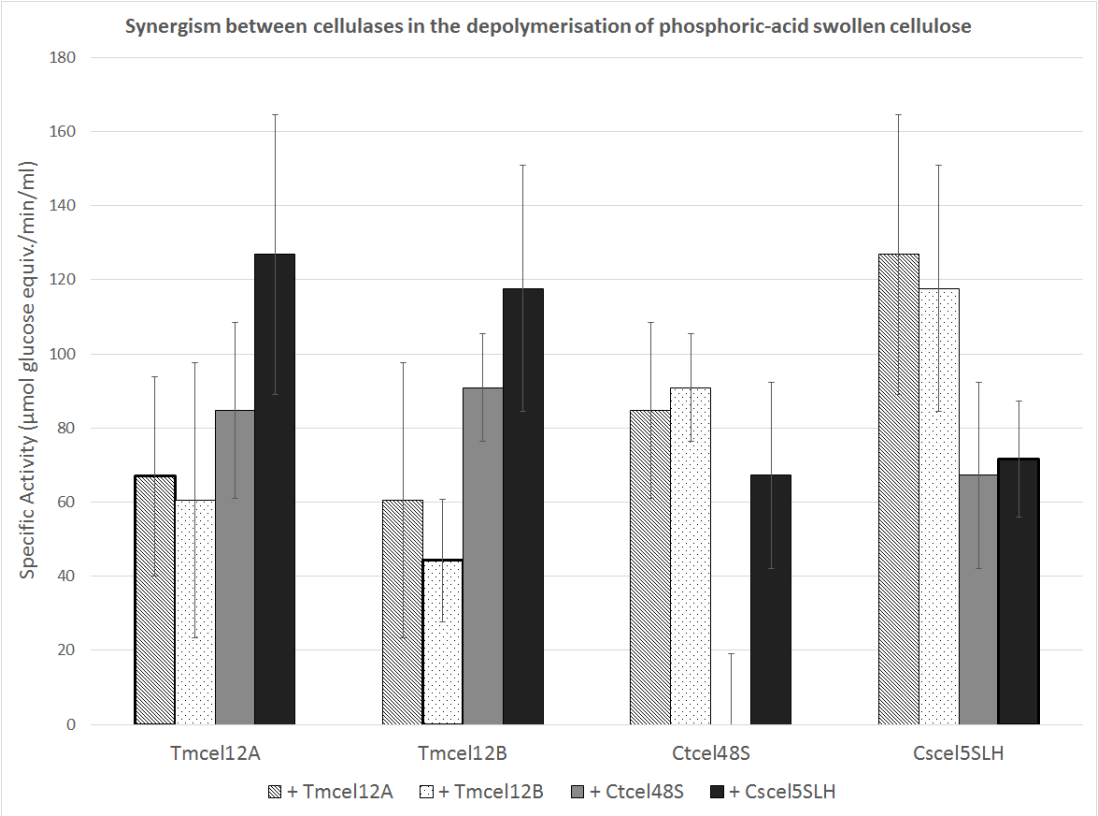


Figure 46 – Individual and combinatorial PASC-degrading specific activities of supernatants from cellulase-secreting *G. thermoglucosidasius* TM242 strains. Supernatants were sampled from 10 ml cultures of exoglucanase Ctcel48S-, and endoglucanase *Tm*cel12A-, *Tm*cel12B- and *Cs*celSLH-secreting *G. thermoglucosidasius* TM242 cultures, grown aerobically in 50 ml centrifuge tubes. These supernatants were filtered using 0.22 µm syringe filters and subsequently mixed in multiple combinations in a 1:1 equivolume ratio. These cellulase mixtures were tested for PASC-degrading activity. For measuring individual cellulase activities, 0.1 M KH_2PO_4 buffer (pH 7.0) was added in place of a paired cellulase supernatant. Synergy factors are calculated by dividing the specific activity of the cellulases in combination by the sum of their individual specific activities (Equation 2).

Equation 2

$$\text{Synergy Factor} = \frac{Specific\ Activity(Enzyme\ A + Enzyme\ B)}{Specific\ Activity(Enzyme\ A) + Specific\ Acitivity(Enzyme\ B)}$$

As shown in **Figure 46**, only the combination of *Tm*cel12B and *Ct*cel48S exhibited a significant level of synergism relative to their individual characterised specific activities. The exoglucanase *Ct*cel48S, in the absence of a paired glycosyl hydrolase, exhibited no detectable activity on PASC, whereas the endoglucanase *Tm*cel12B exhibited a moderate activity of 44.2 U/ml in comparison to the endoglucanases *Cs*cel5SLH and *Tm*cel12A. However, the

combination of the two resulted in a specific activity of 91 U/ml, translating to a synergy factor of 2.05.

Intriguingly, although the additive specific activities of Tmcel12A (67.0 U/ml) and Tmcel12B (44.2 U/ml) should result in a combined specific activity of 111.2 U/ml, the measured specific activity of this *T. maritima* endoglucanase combination was markedly lower at 60.5 U/ml. This may point towards a potential degree of antagonism between *Tm*cel12A and *Tm*cel12B.

Therefore, the combination of *Tm*cel12B- and *Cs*cel48S-secreting *G. thermoglucosidasius* TM242 in co-culture may potentially serve as an efficacious pairing on the degradation of PASC *in vivo*.

5.4 Investigation of expansin-like protein enhancement of cellulase activities.

As discussed in **section 4.3**, there is increasing interest in the exploitation of expansin-like proteins that have been shown to enhance the activity of cellulases in depolymerisation studies. Although there have been no reports of an expansin-like protein with thermal activity at the elevated growth temperatures of *G. thermoglucosidasius* TM242, this study identified a distant homolog of the well-characterised *B. subtilis* YoaJ expansin-like protein in the genome of *C. clariflavum* DSM24705, henceforth referred to as *Cc*YoaJ, which was subsequently cloned downstream of the constitutive pUP2n38 promoter and the *G. thermoglucosidasius* C56-YS93 xylanase signal peptide.

To investigate the potential enhancement of cellulase activity by *Cc*YoaJ, supernatants were isolated at a culture OD_{600nm} of 2 AU from the *Cc*YoaJ-secreting *G. thermoglucosidasius* TM242 cultures and the cellulase-secreting *G. thermoglucosidasius* TM242 cultures characterised in **section 5.3.** Cellulase-containing supernatants were then assayed for their specific activity in the depolymerisation of PASC in the presence and absence of *Cc*YoaJ-containing supernatant.

The addition of *Cc*YoaJ significantly increased the PASC-depolymerising activities of the three endoglucanases and the exoglucanase **(Figure 47)**, although measured *Cc*YoaJ activity on PASC was minimal. In regards to the endoglucanases, the synergistic effect of *Cc*YoaJ addition was greater for *Tm*cel12B (194% increase), than for *Tm*cel12A (58% increase) and *Cs*cel5SLH (55% increase). This was particularly notable since *Tm*cel12B exhibited the lowest PASC-

degrading specific activities from the panel of endoglucanases in the absence of *Cc*YoaJ (44.2 U/ml).

Intriguingly, the highest level of synergism was experienced with the exoglucanase *Ct*cel48S, which exhibited no detectable PASC-depolymerising activity, presumably due to the absence of a carbohydrate binding module. However, the addition of *Cc*YoaJ resulted in the highest level of PASC-degrading activity (140.7 U/ml), indicating that the expansin-like protein may facilitate a similar synergistic function to that of the carbohydrate binding module (CBM3) of the *C. thermocellum* CipA scaffoldin, which natively improves *Ct*cel48S activity.

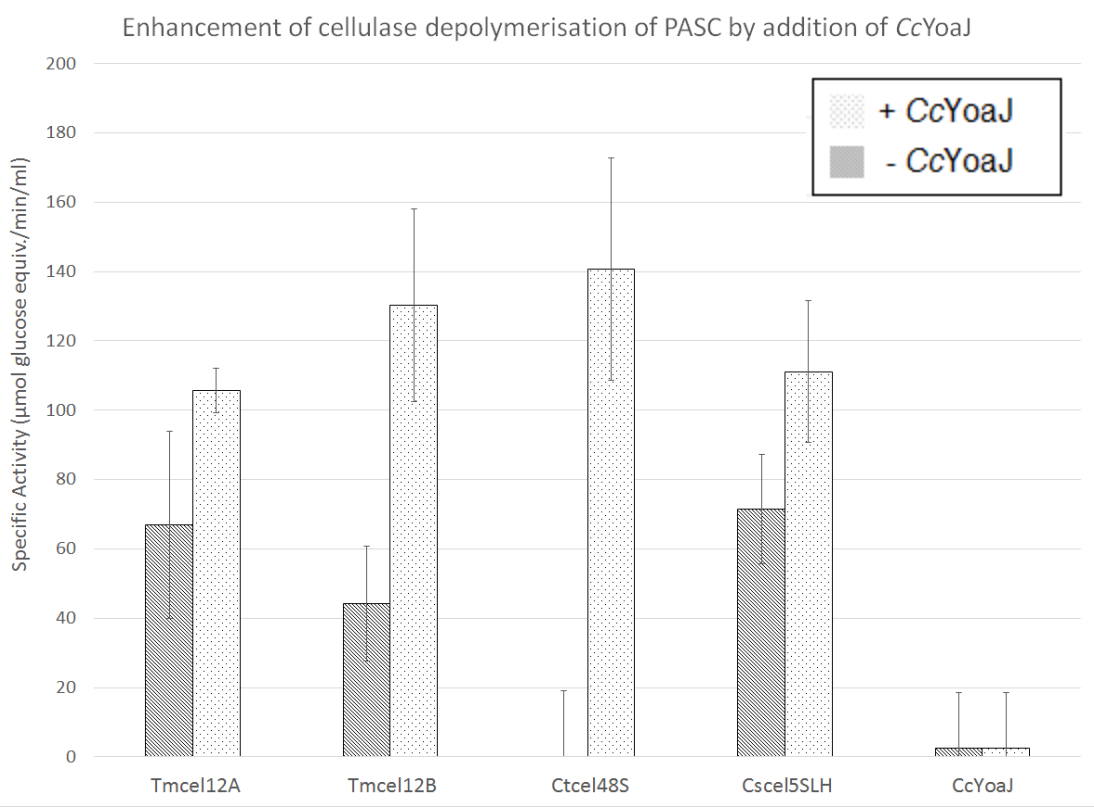


Figure 47 – *Cc*YoaJ expansin-like protein enhancement of PASC-degrading specific activities of supernatants from cellulase-secreting *G. thermoglucosidasius* TM242 strains. Supernatants were sampled from 10 ml cultures of exoglucanase *Ct*cel48S-, and endoglucanase *Tm*cel12A-, *Tm*cel12B- and *Cs*celSLH-secreting *G. thermoglucosidasius* TM242 cultures, grown aerobically in 50 ml centrifuge tubes. These supernatants were filtered using 0.22 µm syringe filters and subsequently mixed in a 1:1 equivolume ratio with either supernatant containing *Cc*YoaJ [+CcYoaJ] or 1M KH_2PO_4 buffer (pH 7.0) [-CcYoaJ] and tested for the release of sugars from PASC.

Although more analysis is required of synergistic activities *in vivo*, this preliminary exhibition of *Cc*YoaJ-mediated enhancement of cellulase activity was promising.

5.5 *Geobacillus thermoglucosidasius* cultures on amorphous carboxymethylcellulose

During the CBP process, multiple polysaccharides will be present for depolymerisation, including crystalline cellulose, amorphous cellulose and xylan. Although the most readily utilisable polysaccharide is likely to be xylan, amorphous cellulose makes up a significant proportion of lignocellulosic biomass, and has reported to be 20 times more accessible than crystalline cellulose (Zhang and Lynd, 2005). Therefore, a potentially effective CBP strategy is the immediate growth of the engineered CBP strain on the less recalcitrant xylan and amorphous cellulose via the activity of highly active endoglucanases and xylanases, while lower-activity exoglucanases continue to decrystallise the more recalcitrant crystalline cellulose.

Therefore, it was important to ascertain the endoglucanases with the combination of properties to facilitate the primary objective of maximum attainable *G. thermoglucosidasius* TM242 biomass on amorphous cellulose. Many factors that were likely to influence this include the specific activities, thermal stabilities and co-factor dependence of the endoglucanases. To investigate this, co-cultures of endoglucanase-expressing *G. thermoglucosidasius* TM242 strains were grown in ASM mineral salts medium supplemented with 1.25% carboxymethylcellulose (CMC) as its sole carbon source.

The maximum level of growth sustained on amorphous CMC was exhibited by the co-culture of *G. thermoglucosidasius* TM242 strains expressing *Ct*cel9I, *Tm*cel12A and *Tm*cel12B separately, which reached a culture OD_{600nm} of 0.67 **(Figure 48 and Figure 49)**. The single culture of the *Ct*cel9I-expressing strain reached a culture OD_{600nm} of 0.8 before immediately reducing thereafter. This suggests that *Ct*cel9I could be a major player in the depolymerisation of amorphous cellulose due to its processive activity. However, this processive activity was insufficient to maintain sustained growth on CMC. Since *Ct*cel9I has been shown to have lower specific activity on methylated cellulose compared to natural unmethylated cellulose, it probably released longer chain oligosaccharides from CMC (Gilad et al., 2003). The *T. maritima* endoglucanases *Tm*cel12A and *Tm*cel12B have high specific activities on CMC, and therefore were likely to have facilitated the subsequent digestion of these longer chain oligosaccharides into transportable sugars in the triple co-culture, resulting in the increased sustainability of growth on CMC.

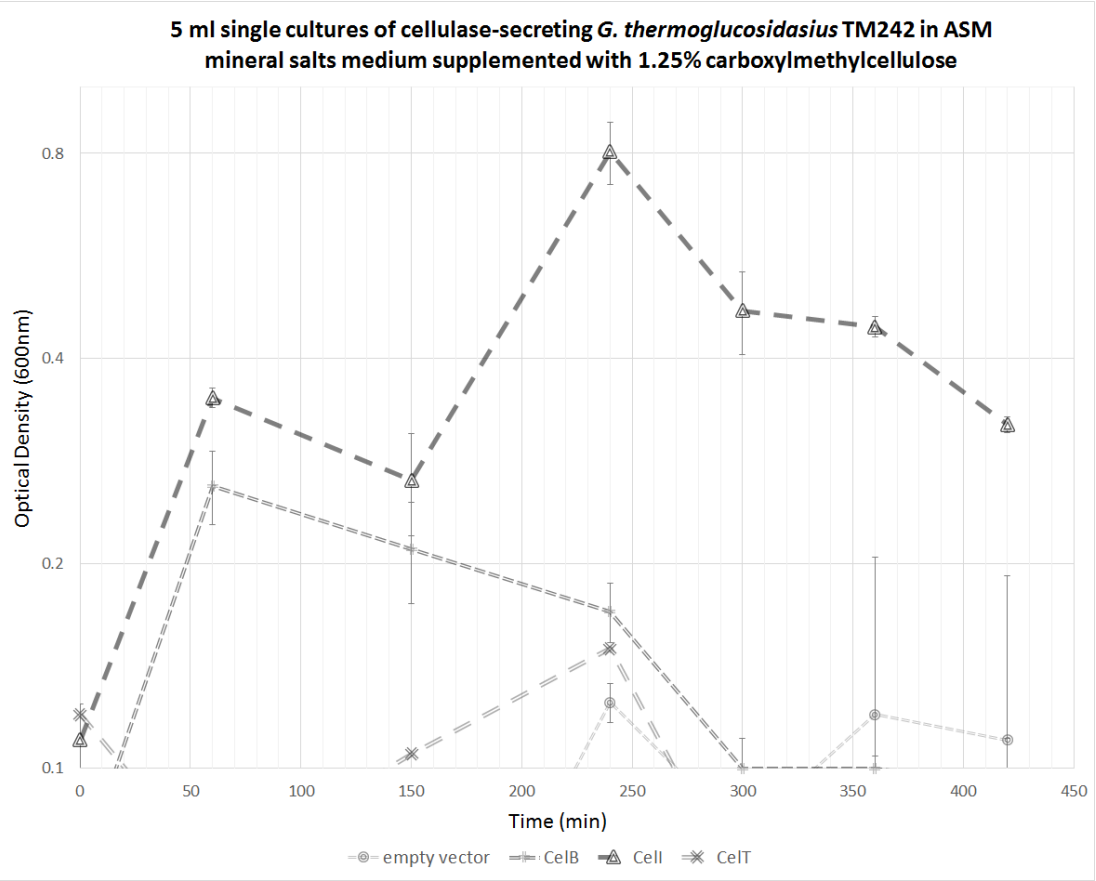


Figure 48 – Measured growth of cellulase-expression *G. thermoglucosidasius* TM242 single cultures on amorphous CMC. Culture OD_{600nm} was measured from 5 ml aerobic cultures of *Ct*cel9I, *Tm*cel12A, and *Tm*cel12B-secreting *G. thermoglucosidasius* TM242 strains grown on ASM minimal salts media with 1.25% CMC as a sole carbon source in 50 ml centrifuge tubes. Empty vector: pUCGRPLS_SigPep1_GoldenGate-expressing *G. thermoglucosidasius* TM242. CelT: pUCGRPLS_SigPep1_*Tm*cel12A-expressing *G. thermoglucosidasius* TM242. CelB: pUCGRPLS_SigPep1_ *Tm*cel12B-expressing *G. thermoglucosidasius* TM242. CelI: pUCGRPLS_SigPep1_ *Ct*cel9I-expressing strains.

Although the triple co-culture provided the highest level of sustained growth, a co-culture of *Ct*cel9I and *Tm*cel12A provided similar levels of growth. This was unexpected, since the characterised specific activity of *Tm*cel12B on CMC was markedly greater than that of *Tm*cel12A. However, although the co-culture of *Tm*cel12B with *Ctcel9I* resulted in culture OD_{600nm} of 0.7, this could also not be sustained. In addition, *Tm*cel12A is not a native extracellular endoglucanase, due to the absence of a native N-terminal signal peptide, and therefore is not likely to natively experience a polysaccharide substrate. *Tm*cel12A is probably adapted to facilitate the intracellular depolymerisation of transported oligosaccharides that require further size reduction for efficient glucose release by

intracellular *B*-glucosidases in *T. maritima.* This activity, when co-cultured with *Ct*cel9I, which may release similar length oligosaccharides to those naturally experienced by *Tm*cel12A, may reduce the size of these *Ct*cel9I products for transport into *G. thermoglucosidasius* TM242. Conversely, *Tm*cel12B may favour the depolymerisation of longer chain oligosaccharides or polysaccharides as it is naturally secreted from *T. maritima*.

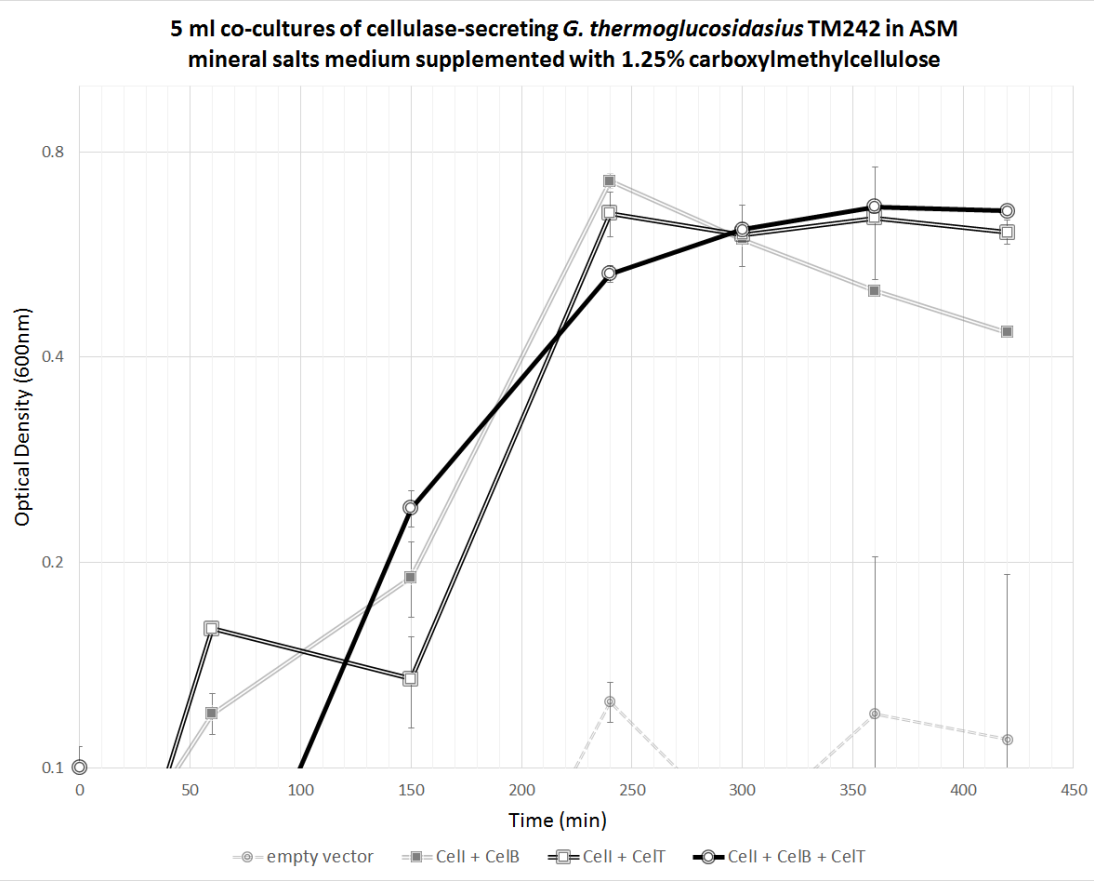


Figure 49 – Measured growth of cellulase-expression *G. thermoglucosidasius* TM242 co-cultures on amorphous carboxymethylcellulose. Culture OD_{600nm} was measured from 5 ml aerobic double- and triple co-cultures of *Ct*cel9I, *Tm*cel12A, and *Tm*cel12B-secreting *G. thermoglucosidasius* TM242 strains grown on ASM minimal salts media with 1.25% CMC as a sole carbon source in 50 ml centrifuge tubes. Co-culture inoculums were pre-mixed and kept constant in volume. Empty vector: pUCGRPLS_SigPep1_GoldenGate-expressing strains. CelT: pUCGRPLS_SigPep1_*Tm*cel12A-expressing strains. CelB: pUCGRPLS_SigPep1_ *Tm*cel12B-expressing strains. CelI: pUCGRPLS_SigPep1_ *Ct*cel9I-expressing strains.

Nevertheless, the triple co-culture experiment was repeated in duplicate flask cultures and compared to duplicate cultures of the *Ct*cel9I-secreting *G. thermoglucosidasius* TM242 strain. As shown in **Figure 50**, the measured OD_{600nm} of cultures did not reach the heights of the closed-culture experiments. This may be due to many different considerations, including

differences in oxygen saturation, shaking strength, and inoculum-culture ratios between the culture experiments.

Nevertheless, the triple co-culture of *Ct*cel9I*, Tm*cel12A *and Tm*cel12B exhibited the highest achieved culture OD_{600nm} of 0.25 AUs **(Figure 50)**, as expected. Given the polymeric nature of CMC as the sole carbon source, the rate of growth was impressive given the preliminary nature of the cellulase system.

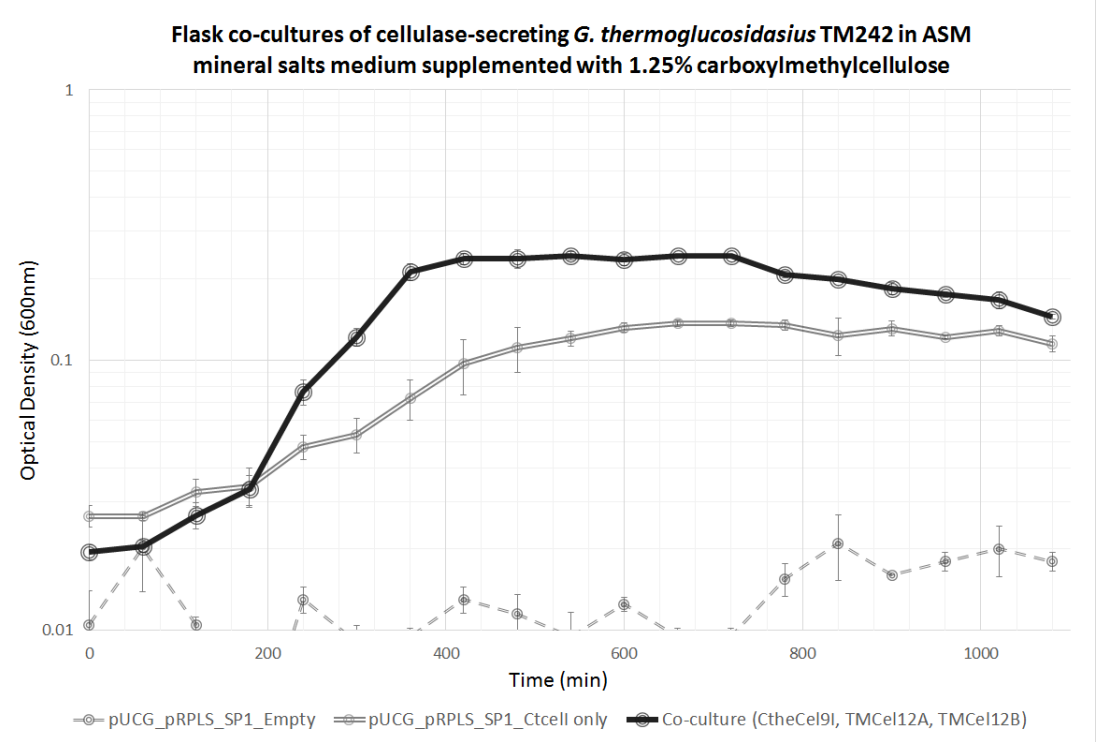


Figure 50 - Measured growth of cellulase-expression *G. thermoglucosidasius* TM242 co-cultures on amorphous carboxymethylcellulose in 50 ml cultures. Culture OD_{600nm} was measured from 5 ml aerobic double- and triple co-cultures of *Ct*cel9I, *Tm*cel12A, and *Tm*cel12B-secreting *G. thermoglucosidasius* TM242 strains grown on ASM minimal salts media with 1.25% CMC as a sole carbon source in 50 ml centrifuge tubes. Co-culture inoculums were pre-mixed and kept constant in volume.

Intriguingly, the long-term sustainability of the culture OD_{600nm}, which only reduced by 0.09 AUs in 11 hours, may support the hypothesis that the extracellular cellulolytic activity conveyed by the secreted heterologous cellulases eventually reached an equilibrium with the cellular energy requirements for increased growth. Essentially, increased culture growth is dependent on resources required for growth, including a carbon source, being greater than the levels required for cellular maintenance energy. The depolymerisation of CMC, which

releases the oligomeric sugar carbon source, is dependent on the overall cellulolytic activity produced by the cellulase-secreting *G. thermoglucosidasius* TM242 strains. This overall cellulolytic activity is, in turn, a function of the levels of extracellular cellulases (cellulolytic capacity), which is dependent on the expression and secretion levels of cellulases, and cellulolytic productivity, which is dependent on the specific activity of the secreted cellulases. Therefore, once the overall extracellular cellulolytic activity reached a level solely sufficient for providing sugars for cellular maintenance energy, but insufficient for further culture growth, an equilibrium was reached.

One potential argument against this hypothesis is that polysaccharides often have branched chains and crystalline regions that reduce the accessibility of cellulases to the substrate, and therefore the insufficiency of cellulolytic activity for culture growth is a consequence of this. Although this is likely the case for crystalline polysaccharide substrates, such as Avicel, endoglucanases have full accessibility to the pure, linear, and amorphous CMC substrate in this experiment.

Further investigation of this potential equilibrium state may provide intriguing findings about the biological economics of engineered CBP systems.

5.7 Cellular growth measurements of cultures with insoluble substrates.

A major objective in the early stages of an envisioned consolidated process is the effective growth of *G. thermoglucosidasius* TM242 cells for high fermentation product yield once anaerobic conditions are set. When the components of the growth medium are all soluble, the concentration of cells can be simply analysed by measured OD_{600}. However, this is not feasible when the culture medium contains insoluble substrates, such as Avicel, PASC, and complex lignocellulosic biomass, which similarly scatter light at 600 nm. Moreover, the use of OD_{600nm} measurements for analysing an increase in light scattering with increased cell concentration is problematic since the potential increase in cell concentration through sugar release is likely accompanied by a decrease in the crystallinity of cellulose, and therefore a decrease in light scattering by cellulosic substrates due to an increase in their solubility.

To circumvent this issue, the use of the Bradford Assay for protein concentration was investigated as a measurement of increased cell concentration. The rationale behind this approach is that an increase in bacterial cell mass results in an increase in protein concentration, which is detected by the Bradford Assay. The Bio-Rad Protein Microassay,

based on the Bradford Assay, was used to analyse increasing concentrations of *G. thermoglucosidasius* TM242 cells resuspended in ASM mineral salts medium supplemented with Avicel as its sole carbon source.

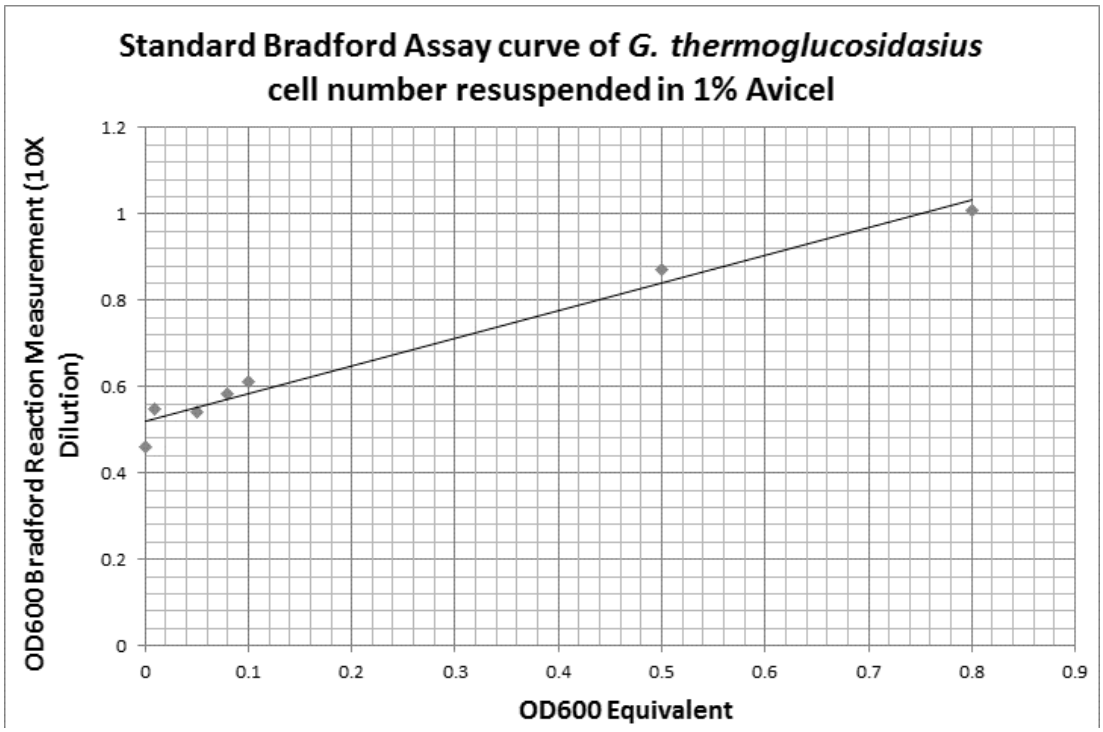


Figure 51 – Bradford Assay standard curve measuring increasing *G. thermoglucosidasius* TM242 cells in the presence of insoluble Avicel. *G. thermoglucosidasius* TM242 cells were resuspended in 2X ASM mineral salts medium, with no added carbon source, to produce cell dilutions that measured an OD_{600nm} of 1.6, 1.0, 0.2, 0.16, 0.1 and 0.02 AUs. 2% w/v Avicel was subsequently added to each cell standard to reduce their culture OD_{600nm} equivalents to 0.8, 0.5, 0.1, 0.08, 0.05 and 0.01 AUs. The Bradford processing reaction was incubated for 10 minutes at room temperature.

As shown in **Figure 51**, a linear increase in light scattering was detected at 600nm post-processing of the sample with the Protein Assay reagent, indicating the detection of increasing proteinaceous cell mass. This method can facilitate the measurement of potential cell growth on Avicel and other insoluble substrates. However, since the maximum limit of accurate measurement of most spectrophotometers is 1 absorbance unit (AU), the measurement of these cultures once an OD_{600nm} equivalent greater than 0.8 have been reached, will require sample dilutions to be made.

In conclusion, the cellulase-secreting *G. thermoglucosidasius* TM242 strains have successfully exhibited extracellular cellulolytic activity and growth on amorphous cellulose. The next

objective for this stage of CBP engineering should be the growth of a co-culture of the most efficacious cellulose-degrading strains on PASC and more recalcitrant substrates. However, it is important to emphasise that the attempted growth of *G. thermoglucosidasius* TM242 on pure cellulose substrates was purely for proof-of-concept, and it is unlikely that pure cellulose will be sole source of carbon in an industrial CBP context. On the contrary, real lignocellulosic biomass substrates are an amalgamation of recalcitrant crystalline cellulose and less recalcitrant amorphous cellulose and hemicellulosic xylan. These latter substrates can serve to increase the growth potential of a cellulase-secreting *G. thermoglucosidasius* TM242 co-culture, as the exoglucanases simultaneously function to decrystallise the recalcitrant crystalline cellulose.

Chapter 6: Analysis and optimisation of protein secretion in *Geobacillus* spp.

Although the universal use of the *G. thermoglucosidasius* C56-YS93 XynA signal peptide had been sufficient for the secretion of heterologous glycosyl hydrolases, further optimisation of glycosyl hydrolase secretion rates may further improve the engineered cellulolytic capabilities of *G. thermoglucosidasius* TM242. However, in order to optimise the secretion of heterologous proteins in *Geobacillus* spp, a greater understanding of natively secreted proteins, often referred to as the secretome, of *Geobacillus* spp. was of paramount use.

6.1 Genomic analysis of *G. thermoglucosidasius* TM242 lignocellulolytic capabilities

Secretion systems in *G. thermoglucosidasius* TM242

In order to improve the secretion capabilities of *G. thermoglucosidasius* TM242, it was important to investigate the putative secretion systems coded for by its sequenced genome. As discussed in **Section 1.10**, the major secretion pathway for protein exportation in bacterial organisms is the Sec pathway. The Sec pathway is subdivided between co-translational secretion of proteins and post-translational secretion of proteins mediated by the recognition of N-terminal signal peptides by either the Secretion Recognition Particle (SRP) protein or SecB/Trigger Factors (TFs), respectively.

As shown in **Table 15**, and illustrated on **Figure 52**, protein homology analysis revealed that *G. thermoglucosidasius* TM242 encodes ORFs with significant sequence identity to most components of the Sec Pathway. Although the presence of a SecB homolog was not identified, this is not an indication that only co-translational protein secretion is likely present. Nevertheless, native expression of secreted proteins, and consequently their co-translational secretion, is expected to be at lower rates than those expressed by the strong synthetic promoters, and therefore sufficient for native co-translational secretion. However, with stronger expression rates, there is a greater likelihood for insufficient secretion rates and, consequently, misfolded proteins that require degradation. This may lead to increased metabolic burden at the expense of potential secretion rates.

The undetected homology between *G. thermoglucosidasius* and *B. subtilis* SRP (Ffh), SecA and SecY components, which provide major functions to Sec-dependent secretion, was surprising, given that homology analysis provided positive results for the minor components of the pathway. Only FtsY, the receptor for the SRP-preprotein complex, showed strong homolog to its *B. subtilis* homolog, with 82% sequence identity spread over a 100% query coverage. However, given the thermophilic nature of *G. thermoglucosidasius* TM242 relative to *B. subtilis* 168, the functional nature of SRP (Ffh), SecA and SecY may have resulted in greater mutational pressures for these components.

Table 15 – *B. subtilis* 168 and *G. thermoglucosidasius* TM242 Sec Pathway component homologs. *G. thermoglucosidasius* TM242 ORFs analysed for >60% sequence identity to characterised Sec pathway components present and absent in *B. subtilis* 168. * denotes an ORF annotated by domain-structure, but with no detectable homology to *B. subtilis* counterpart.

Protein Name	*B. subtilis* 168 Accession	TM242 homolog	Protein Function
SRP54 (Ffh)	NP_389480	**peg0823***	Recognition of N-terminal signal sequence.
SecA	NP_391410	**peg3370*/ peg3498***	Cell membrane associated ATPase.
SecB			Chaperone for fully-translated proteins
SecM			Secretion monitor, involved in secretion-responsive control of SecA translation.
SecY	NP_388017	**peg3669***	Translocase transmembrane subunit.
SecE	NP_387981	**peg3705**	Translocase transmembrane subunit.
SecG	NP_391243	**peg3311**	Translocase transmembrane subunit.
SecD/F	NP_390643	**peg2786**	Required to maintain a proton motive force.
YajC	NP_390648	**peg2790**	Involved in stabilising and regulating SecYEG.
YidC	NP_390269	**peg1495/ peg3805**	Required for insertion and assembly of inner membrane proteins.
FtsY	NP_389477	**peg0821**	Acts as a receptor for SRP-Secretion complex.

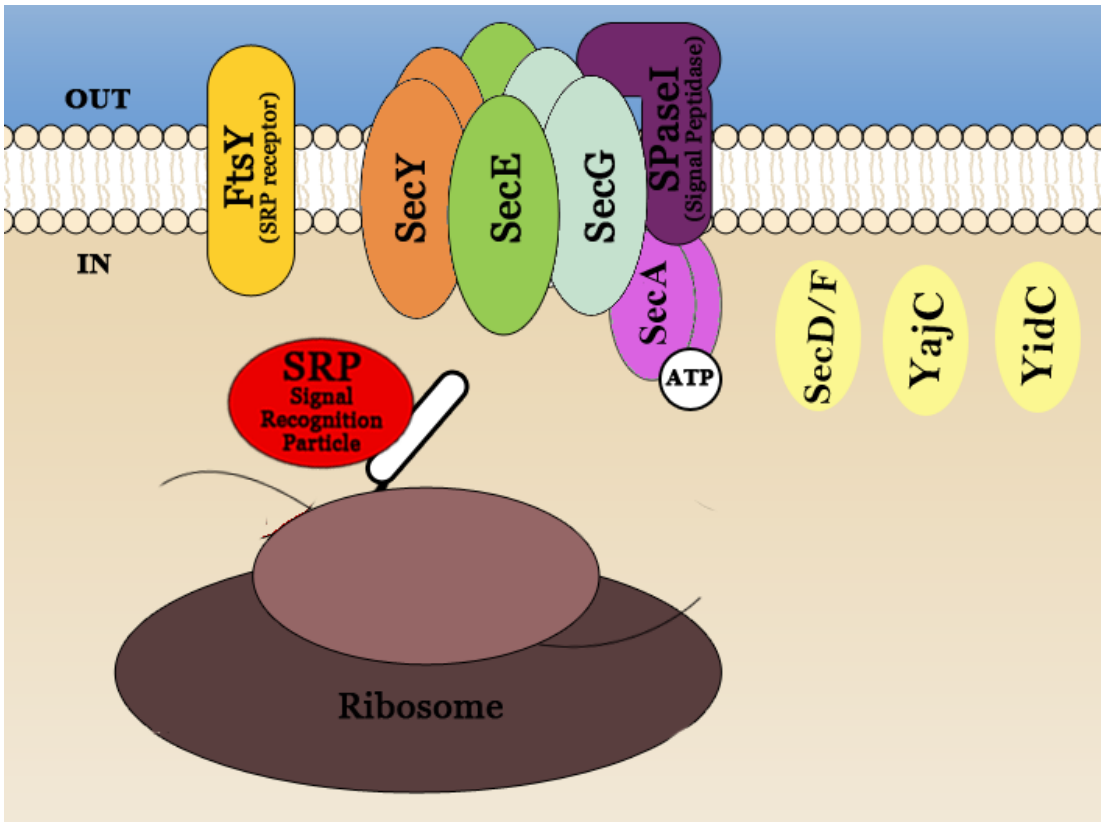


Figure 52 – Schematic of the identified components of the *G. thermoglucosidasius* TM242 Sec Pathway.

Twin Arginine Pathway

An increasingly studied secondary secretion pathway, termed the Twin-Arginine Pathway, facilitates the post-translational export of proteins in their folded native state. This is particularly important for proteins that contain co-factors and chromophores, which do not get co-secreted by the Sec Pathway, but are required for protein function. As shown in **Table 16** and annotated on **Figure 53**, protein homology analysis revealed that *G. thermoglucosidasius* TM242 encodes ORFs with significantly high sequence identity to TatA and TatC components of the Sec Pathway, but not to TatB. However, the absence of TatB is common in Gram-positive organisms, including *B. subtilis*, that are able to substitute for either TatA or TatB to form TatAC systems (Goosens *et al.*, 2015). Moreover, the presence of more than one copy of TatA provides further evidence for its importance and potential dual –functionality, as is the case in *G. thermoglucosidasius* TM242 (Barnett *et al.*, 2008). In fact, *B. subtilis* has been characterised to contain two distinct Tat complexes, one ~230 kDa TatAdCd complex that is significantly smaller than the analogous *E. coli* TatABC complex (~370 kDa), and a separate ~270 kDa TatAd complex, whose singular nature of the *B.*

subtilis TatAd complex suggests that discrete TatAC and TatA complexes may form a single form of translocon.

Table 16 - *B. subtilis* 168 and *G. thermoglucosidasius* TM242 Tat Pathway component homologs. *G. thermoglucosidasius* TM242 ORFs analysed for > 60% sequence identity to characterised Sec pathway components present and absent in *B. subtilis* 168. Although Tatd, TatAy and TatAc ORFs returned as homologs to more than one TatA *B. subtilis* paralog, they can be clearly defined by the *B. subtilis* TatA component of highest sequence identity.

Protein Name	*B. subtilis* 168 Accession	TM242 homolog	Protein Function in *B. subtilis*
TatAd	NP_388145	**peg1551**	Complex expressed only under conditions of phosphate limitations
TatCd	NP_388146	**peg1552**	
TatAy	NP_388479	**peg2089**	Complex expressed constitutively over many tested conditions and exports more substrates
TatCy	NP_388480	**peg4120**	
TatAc	NP_389654	**peg4119**	Complements TatAy via pore formation, but unable to functionally replace TatAy
TatB			Cell membrane associated ATPase

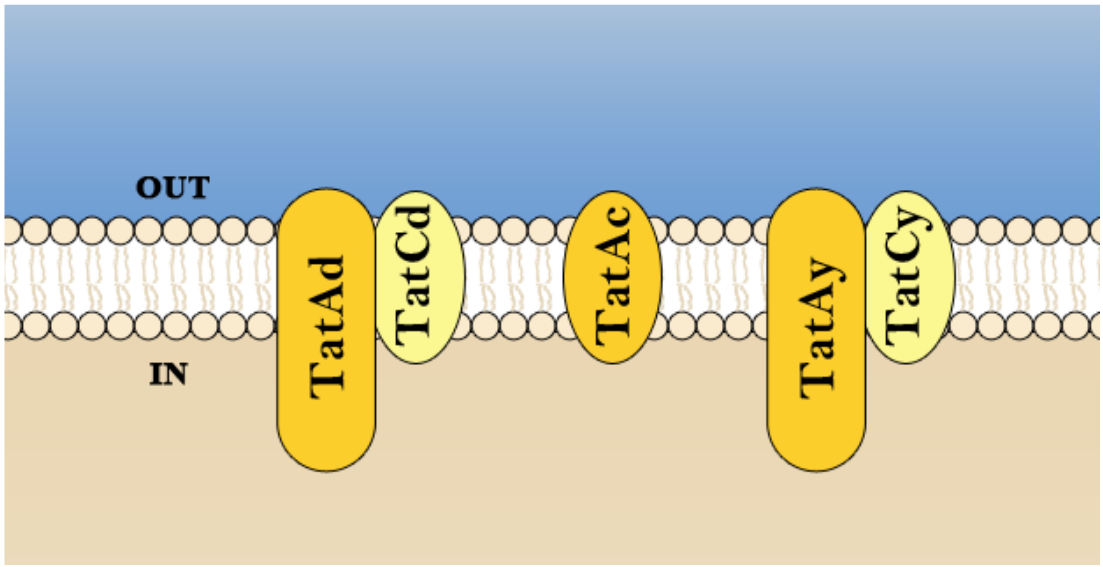


Figure 53 – Schematic of the identified components of the *G. thermoglucosidasius* TM242 Tat Pathways

The Tat pathway presented an interesting alternative to Sec-dependent secretion, and will be required in the future for the secretion of chromophore-containing proteins**.** However, since there was no requirement for the secretion of glycosyl hydrolases in their native form,

the following identification of signal peptides and library creation focused on utilising the Sec Pathway.

6.2 Bioinformatic prediction of signal peptide library

Different prediction servers

The proposed method of improving the efficiency of glycosyl hydrolase secretion in *G. thermoglucosidasius* TM242 was the generation of a library of signal peptides that can be screened for the optimal secretion of a particular glycosyl hydrolase. In order to achieve this, a panel of signal peptides that facilitate native protein secretion from *G. thermoglucosidasius* TM242 were identified by analysing and screening the translated genome of *G. thermoglucosidasius* TM242, and 11 other *Geobacillus* spp. genomes, for predicted secretion signal peptides and against membrane proteins using five prediction tools.

Two of these servers focused on the prediction of signal peptides and signal peptidase cleavage sites in open reading frames (ORFs). SignalP 4.1, developed by the Technical University of Denmark, incorporates a combination of several artificial neural networks for its prediction. PrediSi, developed by the Technical University of Braunschweig, is based on a position weight matrix approach that is corrected for amino acid bias present in signal peptides. Although these two prediction servers had different methodologies, they are both commonly utilised prediction servers to predict for the presence and location of signal peptides.

To separate predicted secretion proteins from predicted transmembrane (TM) proteins, which both encode N-terminal signal peptides, two prediction servers were used to predict the presence of transmembrane helices in translated ORFs. TMHMM 2.0, developed by the Technical University of Denmark, predicts transmembrane helices using a Hidden Markov Model (HMM). SOSUISignal, developed by the University of Nagoya, consolidates a hydropathy index, an amphiphilicity index, an index of amino acid charges, and the length of each sequence in the prediction of transmembrane helices.

Although the glycosyl hydrolases expressed in the CBP system were required to be secreted extracellularly in solution, this did not rule out the use of signal peptides from TM proteins for the screening process, since the intramembrane folding of TM proteins is independent of signal peptide targeting. Nevertheless, due to the lack of empirical characterisation of *G.*

thermoglucosidasius TM242 secretion, and the proposed encoding differences from *B. subtilis* 168 as highlighted in **Section 4.1**, signal peptides from predicted secretion proteins were favoured for addition to the proposed *G. thermoglucosidasius* signal peptide library.

LipoP 1.0, also developed by the Technical University of Denmark, had been released for the prediction of lipoproteins (Juncker *et al.*, 2003). However, this server offered further discrimination between lipoprotein signal peptides, Sec-SRP signal peptides, and n-terminal membrane helices, Although the server's Hidden Markov Model was trained on sequences from Gram-negative bacteria only, studies had reported it to perform well on sequences from Gram-positive bacteria (Rahman *et al.*, 2008). Nevertheless, LipoP analysis of translated ORFs was purely used for ratification of signal peptide and transmembrane helix predictions made by the four aforementioned prediction servers.

The stringency of soluble secreted protein prediction between four prediction servers.

Discrepancies between prediction results occurred more frequently than expected. Using a combination of IF and COUNT functions *(Microsoft Excel)*, prediction results from SignalP, PrediSi, TMHMM and SosuiSignal were compared to calculate the number of soluble secreted proteins as predicted by one, two, three or all the prediction servers. An example of such a discrepancy is the *Geoth_0344* gene of *Geobacillus thermoglucosidasius* C56-YS93, which codes for a predicted subtilisin protein [YP_004586469.1]. The identification of three SLH domains present at the carboxyl-terminus of the predicted protein put some weight on the probability of this predicted protease being a soluble secretion protein, as is the case for all S-layer proteins that attach to the cell surface via a post-secretion mechanism. However, the ORF was predicted to contain a 25 amino-acid N-terminal signal peptide by the PrediSi prediction server, but no signal peptide was predicted by SignalP. Similarly, SosuiSignal predicted the presence of one transmembrane-spanning helix, whereas TMHMM did not predict a transmembrane helix.

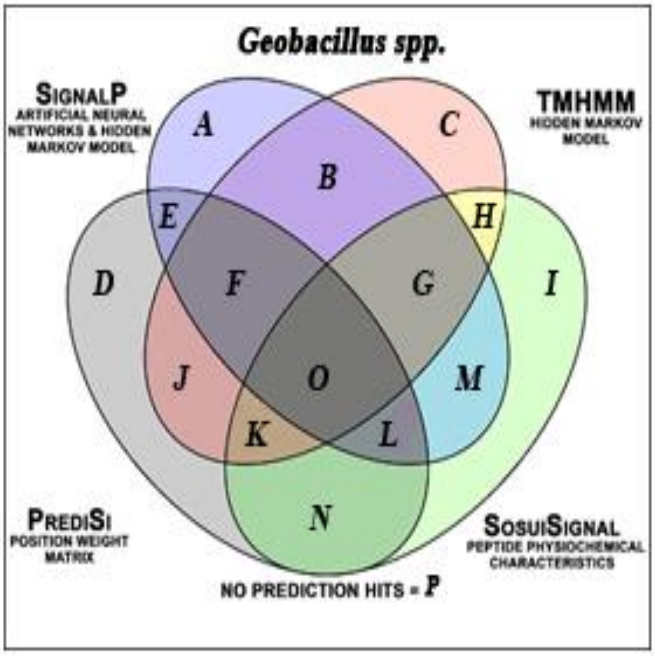


Figure 54 – Schematic of four-set Venn diagram overlaps of positive prediction counts of results from SignalP, PrediSi, TMHMM and SosuiSignal servers. A positive prediction by SignalP and PrediSi is the predicted presence of a signal peptide cleavage site (signal peptide prediction score = 0). A positive prediction by TMHMM and SosuiSignal is the absence of predicted transmembrane helices (TM prediction score = 0).

The results from this comparison study are shown in four-set Venn diagrams **(Figure 54)** to illustrate the numbers of agreements and disagreements between prediction servers of similar purposes, but differing methodologies. Certain overlaps can be interpreted to provide important estimates for the subcellular localisation of proteins in the cytoplasm, membrane and extracellular milieu.

For example, the overlap of most interest is overlap O, which represents the most confident components of the predicted soluble secretome (high-confidence secretion proteins), counts the ORFs predicted by all prediction servers to code for soluble proteins that contain signal peptides. Overlap H represents the ORFs with the highest confidence prediction scores for encoding intracellular soluble proteins, since both TMHMM and SosuiSignal predicted the absence of TM helices, and both SignalP and PrediSi predicted the absence of an N-terminal signal peptide. As expected, this overlap contains the largest count of predicted ORFs. Lastly, overlap E represents the ORFs with the highest confident prediction scores for encoding transmembrane proteins, due to the predicted presence of an encoded N-terminal signal peptide, as predicted by both SignalP and PrediSi, and the presence of TM helices, as predicted by TMHMM and SosuiSignal.

The analysis of both intra- and interspecies variations in prediction counts provided intriguing results. For instance, *Geobacillus* sp. C56-T3 stood out as having the largest secretome, with its prediction count of high-confidence secretion proteins (Overlap E: 50) being 35% greater than the average (37), and 150% greater than the corresponding prediction count for *Geobacillus* sp. WCH70 (Overlap E: 20), which had the smallest predicted secretome **(Figure 56)**. This was not surprising, since *Geobacillus* spp. are likely to have evolved different survival strategies in the various climates they have been isolated from. However, the low count of high-confidence secretion proteins specifically in *Geobacillus* sp. WCH70 was surprising given the discovery of a putative SecB-coding ORF (YP_002951153.1) in its genome, which should more likely be retained in a strain with a large secretome, and therefore high secretion load.

Moreover, although there was little significant intra-species variation in regards to their putative secretion protein and transmembrane protein prediction counts, *G. thermoglucosidasius* TM242 and *G. thermoleovorans* CCB_US5_UF5 were predicted to encode notably more soluble cytoplasmic proteins (overlaps C, H and I) than their counterpart strains **(Figure 55)**. This may be either due to a greater emphasis on intracellular metabolic activities leading to more cytoplasmic protein-coding ORFs, or a function of the high frequency of hypothetical protein ORFs in *Geobacillus* species. A subset of these hypothetical protein annotated ORFs may have been incorrectly identified, and are simply a region of DNA that serendipitously contains a long ORF in a single translation frame. Since these incorrect ORFs are unlikely to contain signal peptides and transmembrane helices, which require specific motifs in amino acid sequence, they are likely to be categorised as soluble cytoplasmic proteins (overlaps C, H and I).

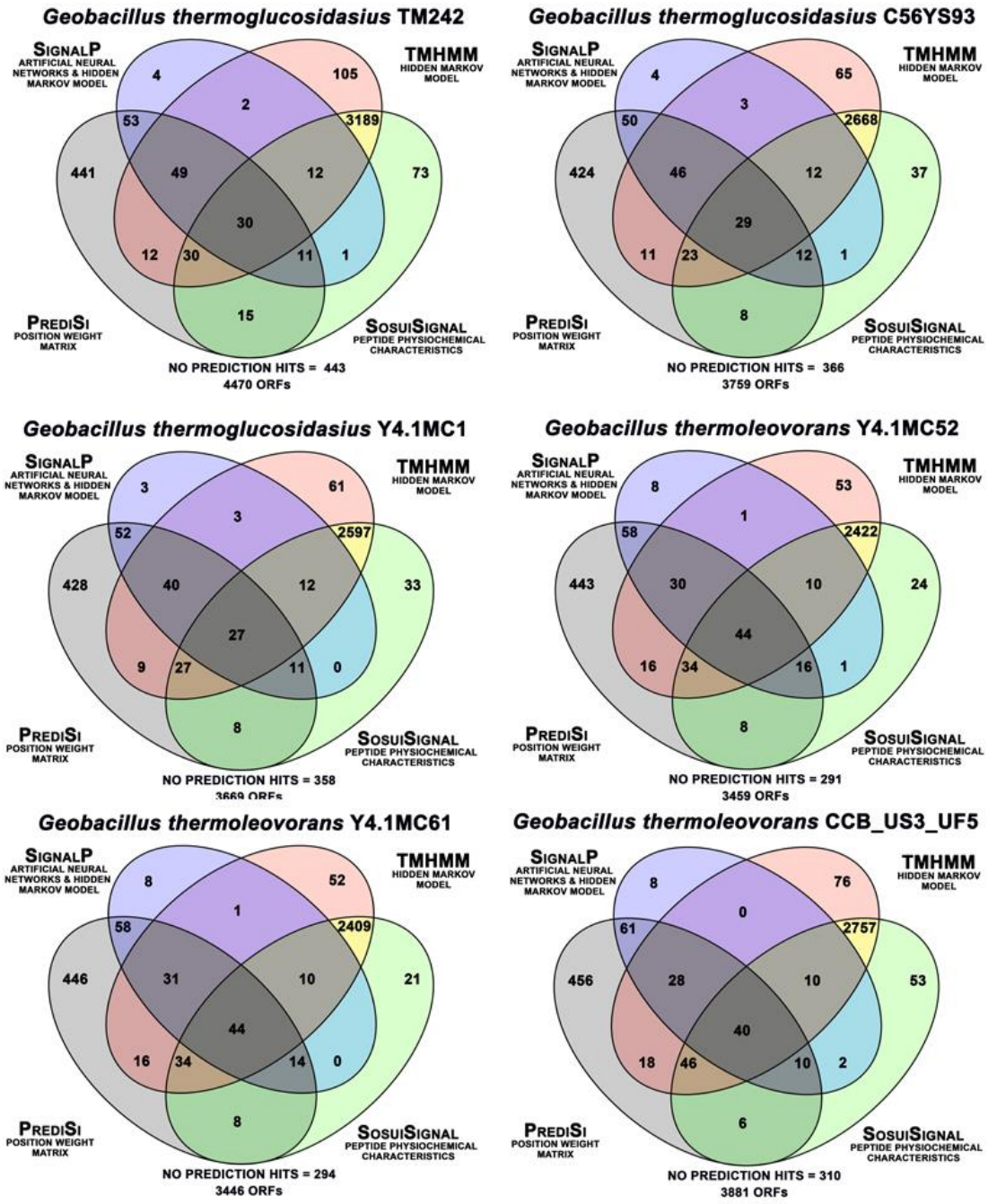


Figure 55 – Four-set Venn Diagrams illustrating discrepancies between predictions by four common signal peptide prediction servers of ORFs from three *G. thermoglucosidasius* strains and three *G. thermoleovorans* strains. A positive prediction by SignalP and PrediSi is the predicted presence of a signal peptide cleavage site (signal peptide prediction score = 0). A positive prediction by TMHMM and SosuiSignal is the absence of predicted transmembrane helices (TM prediction score = 0).

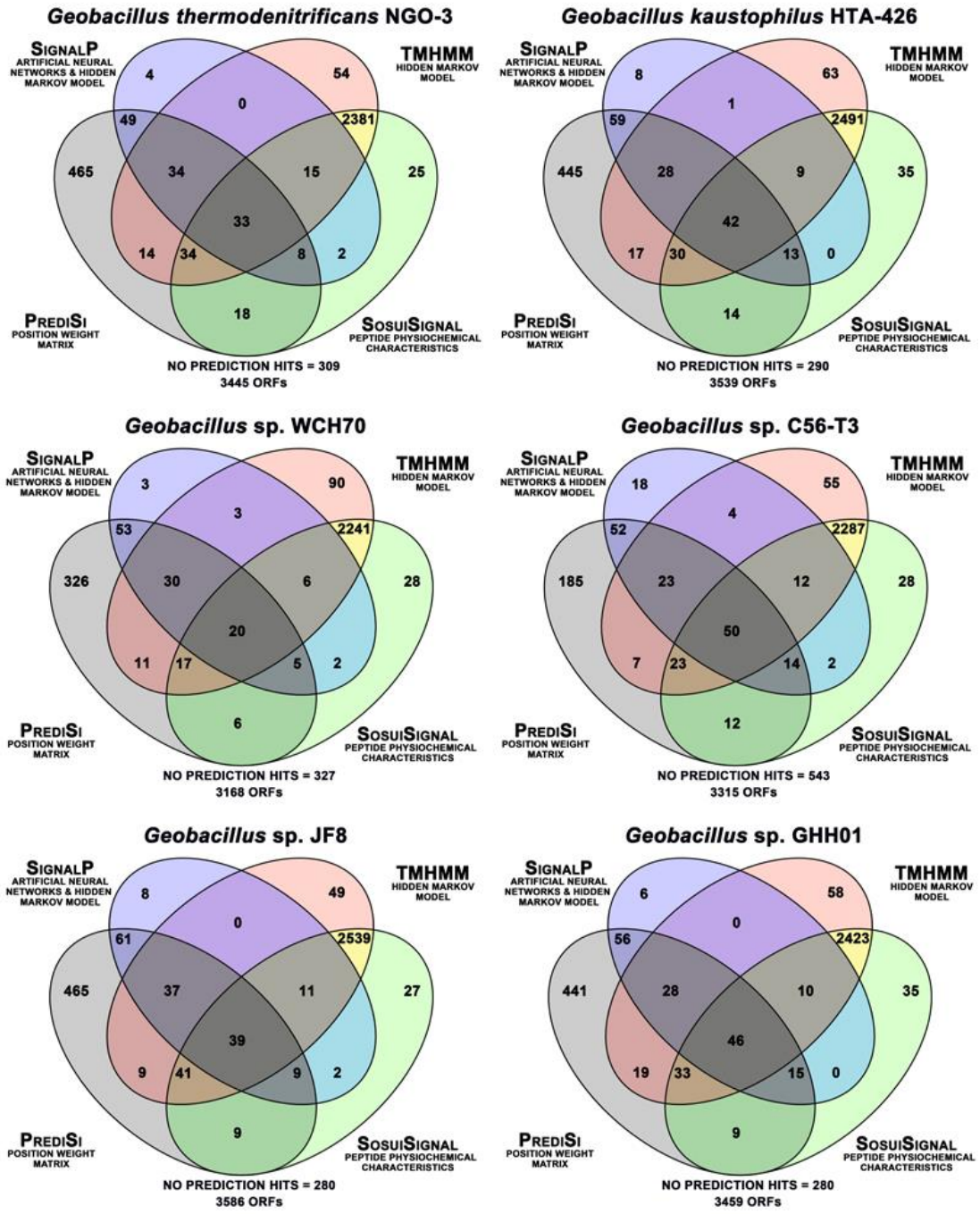


Figure 56 – Four-set Venn Diagrams illustrating discrepancies between predictions by four common signal peptide prediction servers of ORFs from six different *Geobacillus* species. A positive prediction by SignalP and PrediSi is the predicted presence of a signal peptide cleavage site (signal peptide prediction score = 0). A positive prediction by TMHMM and SosuiSignal is the absence of predicted transmembrane helices (TM prediction score = 0).

A consolidated prediction score for signal peptide prediction

The consolidation of the results from the four prediction tools could, albeit *in silico*, serve useful in reducing the impact of discrepancies between the prediction servers. A consensus score was calculated from the four prediction servers by calculating the average of six confidence factors. These were the highest confidence score for signal peptidase cleavage (SignalP: *Cmax*), the mean confidence score for the presence of a signal peptide (SignalP: *Smean*), the maximum confidence score returned for an amino acid in the predicted signal peptide (SignalP: *Smax*), the general signal peptide prediction score returned by PrediSi, and the number of predicted transmembrane helices, predicted by TMHMM and SosuiSignal, subtracting from 1 to generate a bias against signal peptides from TM proteins.

$$\boldsymbol{Consensus\ score} = \frac{\boldsymbol{C} + \boldsymbol{M} + \boldsymbol{S} + \boldsymbol{P} + (\mathbf{1} - \boldsymbol{T}) + (\mathbf{1} - \boldsymbol{H})}{\mathbf{6}}$$

Equation 2 – *Geo*Sec score equation, where *C* is the Cmax value corresponding to the predicted presence of a SPase cleavage site (SignalP). M and S are the Smean and Smax values respectively, corresponding to the predicted presence of signal peptides by analysing amino acid positioning (SignalP). P is the output prediction score for signal peptide cleavage using PrediSi. T and H are the number of predicted transmembrane helices calculated by TMHMM and SosuiSignal respectively.

This *Geobacillus* secretion (Geosec) consensus score should have theoretically reduced the likelihood of false negative and false positive results predicted by the individual prediction servers, by complementing each individual score with a score from a corresponding prediction server that uses a different prediction methodology. It is fair to argue that the consolidation of these scores may have led to a group of open reading frames (ORFs) being incorrectly predicted to not express a signal peptide, although the prediction of one or more prediction servers was accurate. However, the purpose of this scoring system is to isolate a subset of ORFs that are strongly predicted to be soluble and secreted, which can lead to the exploitation of their signal peptides for the purpose of building a signal peptide library.

6.3 Construction of *G. thermoglucosidasius* signal peptide library

Based on the calculated GeoSec scores of 4470 predicted ORFs in the *G. thermoglucosidasius* TM242 genome and 3759 predicted ORFs in the *G. thermoglucosidasius* C56-YS93 genome, 24 signal peptides from the 24 highest scoring ORFs were selected for signal peptide library construction **(Table 17)**. Signal peptide fragments were amplified from *G. thermoglucosidasius* TM242 and C56-YS93 genomic DNA using primers 64-110.

Table 17 – The 24 signal peptides of the Geobacillus spp. signal peptide (SP) library. SPs were categorised by predicted length and the highest GeoSec-scoring ORFs were selected.

Signal Peptide	Strain	Amino Acid Sequence	Size (aa)	GeoSec Score
Signal Peptide 02	11955	MKRMLTGCLLASLLFAFPAMAA	22	0.824
Signal Peptide 03	11955	MRWILAAMLVLSSFFSISASAAA	23	0.807
Signal Peptide 04	C56-YS93	MRMLAAFFVMFFHALFVQAAFAA	23	0.774
Signal Peptide 05	C56-YS93	MKRIWLLAFIAFIYAFPPHANAA	23	0.708
Signal Peptide 06	11955	MKTRWLFLAAALMLMLPTGTLAAQ	24	0.792
Signal Peptide 07	11955	MKGWSKFFICLCLLFAFHLPVQAQ	24	0.773
Signal Peptide 08	11955	MMKKQLYVWLMIVLLLVPWHASAE	24	0.758
Signal Peptide 09	11955	MKVKAVAVVGLFFFFFSPFVVQAS	24	0.751
Signal Peptide 18	11955	MKKWKKTAVSLGLASALVLPSFAQAS	26	0.856
Signal Peptide 26	11955	MKKWKWYLTAFLCFCMVFGFLLPANAK	27	0.764
Signal Peptide 27	11955	MKRRKVMALAAATVLSIGVFPHFADAV	27	0.737
Signal Peptide 28	C56-YS93	MIKKTFAALAVVCCIVSGISSYSTDAA	27	0.721
Signal Peptide 29	11955	MNAVKATIPVLTAATLLLSSATGTYAAA	28	0.792
Signal Peptide 30	11955	MCLRRMRLLLFLFCLSMVVGMVLPVLAA	28	0.782
Signal Peptide 31	11955	MLSFYKKITVILVAVVMLFVPWTSPQAH	28	0.769
Signal Peptide 32	11955	MRRQLVLALLLGGSVFAAGARAEQAEAS	28	0.742
Signal Peptide 33	C56-YS93	MLQSRIIALFLCIVVAIAAQTEFPVFAE	28	0.731
Signal Peptide 38	C56-YS93	MKGRRRLMMFCFPFLCSVLAAMGMTVYAE	29	0.706
Signal Peptide 40	C56-YS93	MAEKRKFLWLLMALLLCVAFGNVPAVAFGA	30	0.819
Signal Peptide 41	11955	MNKTKSYLSFLLSFVLVLSTLGGAGIAQAQ	30	0.793
Signal Peptide 42	11955	MRIGVQIRKFAALLSVLILLVSYAISPAYAA	31	0.843
Signal Peptide 43	11955	MKKWILAMLSVSVLCLLVYFVIGQAENVFAD	31	0.757
Signal Peptide 44	C56-YS93	MKLPKWLRKVLVVTITVCTFGLVTPPASLMAA	32	0.717
Signal Peptide 45	C56-YS93	MRLFKFAFIRMMLICFTFFSVFSPSSLSVMAA	32	0.714

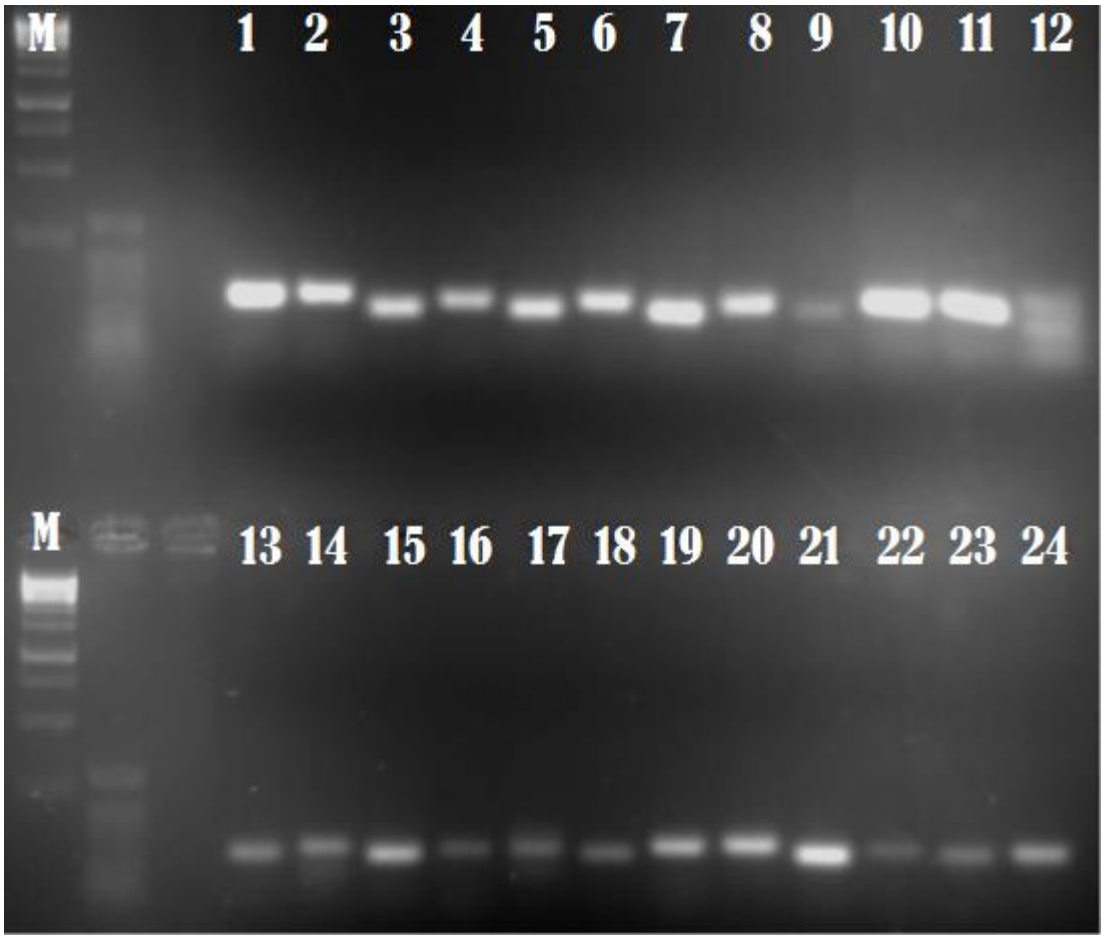


Figure 57 - High-fidelity PCR amplification of the 24 signal peptide components of the *Geobacillus* spp. signal peptide library. A 4% TAE agarose gel of 24 signal peptides, of which 16 were amplified from *G. thermoglucosidasius* TM242 genomic DNA, and the remaining 8 were amplified from *G. thermoglucosidasius* C56-YS93 genomic DNA. To ensure sufficient PCR product for visualisation, the high-fidelity Phusion polymerase was used and the number of amplification cycles was increased to 45.

Since it was important to preserve the signal peptide library for future use, the signal peptide fragments were collectively inserted into a pRPLS-containing pUCGXXX system. This was achieved by the digestion of the amplified signal peptide library fragments with *Cla*I, which cleaves immediately upstream of their ATG start codon at the 5' end of the fragment. This was followed by standard ligation of the pool of signal peptides into pUCG4.8_pRPLS_SigPep1_*Ct*cel48S, which had been digested with *Cla*I and the blunt-end creating enzyme *Sma*I. The *Cla*I overhangs of the pool of signal peptides (76 bp – 106 bp) and the digested vector (5,916 bp) annealed selectively, and the blunt ends generated by *Sma*I digestion were sufficient for ligation of the 3' blunt PCR product ends generated by the amplification of the signal peptides. *E. coli* transformants of this signal peptide library mixture, henceforth referred to as pUCG4.8_pRPLS_SPL1.0, were cultured and stored as glycerol stocks.

6.4 Demonstration of *G. thermoglucosidasius* signal peptide library for the optimisation of *Tm*cel12A secretion.

To demonstrate the applications of the signal peptide library, the endoglucanase *Tm*cel12A was chosen as the preliminary test cellulase for protein secretion optimisation. To generate a pool of *Tm*cel12A-secreting plasmisd constructs, each with a different signal peptide driving the secretion of the cellulase, the plasmid construct pUCG4.8_pRPLS_SigPep1_*Tm*cel12A, which contained the *Tm*cel12A mature protein coding sequence fragment (MCS), and pUCG4.8_pRPLS_SPL1.0, the signal peptide library plasmid mix, were both digested with *Ngo*MIV and *Sac*I. After resolving on a 0.8% gel, the ~5,100 bp pUCG_pRPLS_SPL1.0 fragment and ~850 bp *Tm*cel12A MCS fragment were purified and ligated in a 3:1 vector-to-insert molar ratio.

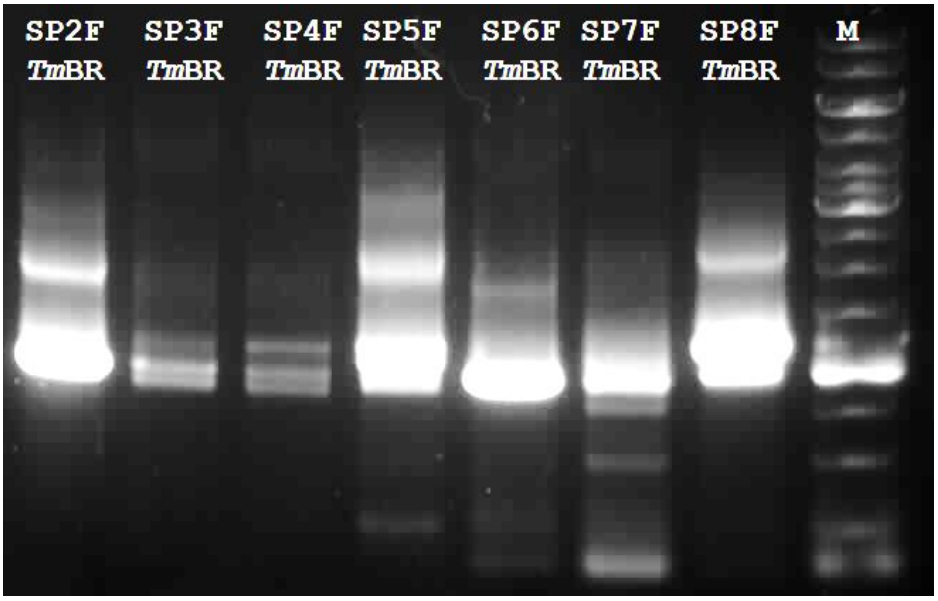


Figure 58 – Diagnostic PCR amplification of a subset of Signal Peptide-*Tm*cel12A combinations
24 forward primers, designed to anneal specifically to the 5' end of a single signal peptide in the signal peptide library, were supplemented to pre-mixed RedTaq polymerase mixture (*Sigma Aldrich*) and a reverse primer annealing specifically to the 3' end of *Tm*cel12A. The first 7 reactions are presented here.

The transformation of *E. coli* Bioblue cells with the pRPLS-expressing *Tm*cel12A signal peptide library resulted in several hundred ampicillin-resistant transformants. The transformant colonies were pooled together, collectively cultured in kanamycin-supplemented rich 2xTY media to eliminate any ampicillin-resistant false-positive transformants, and plasmid DNA was purified. The resulting plasmid library of multiple *G. thermoglucosidasius* signal peptides driving the secretion of the same endoglucanase, was tested for library completion by

diagnostic PCR amplification using signal peptide-specific forward primers and a *Tm*cel12A-specific reverse primer **(Figure 58)**.

Once library completion was verified, the plasmid library was subsequently transformed into *G. thermoglucosidasius* TM242. The availability of the congo-red agar staining for glycosyl hydrolases as a simple metachromatic activity agar assay facilitated rapid semi-qualitative screening for optimal signal peptide-endoglucanase combinations. *G. thermoglucosidasius* TM242 transformants were spotted onto ASM minimal media containing 1% carboxymethylcellulose (CMC) as a sole carbon source, and supplemented with 12.5 µg/ml kanamycin. After incubated growth at 60°C for 16 hours, the dimensions of resultant colonies were measured prior to the addition of 10 ml of 1% Congo Red. The measurement of colony dimensions is important for semi-qualitative comparisons of secretion rates, and since congo red staining leads to the solubilisation of colonies, these measurements had to be taken prior to congo red addition. After 20 minutes of congo-red staining, the plates were washed and destaining by the addition of 1M NaCl.

Zones of clearing were observed surrounding a subset of colonies **(Figure 59)**, and since the sole difference between the colonies is the signal peptide fragment, these size of the zones of clearing could be directly attributed to the strength of *Tm*cel12A secretion by the corresponding colony. The high constrast in secretion rates is visually obvious, with colonies J and V evidently exhibiting high secretion rates of the *Tm*cel12A endoglucanase. Conversely, a faint zone of clearing can be seen surrounding the shaded colony K, indicating suboptimal secretion rates. Moreover, it was observed that colonies grew on the CMC agar that did not exhibit a zone of clearing, including the pUCG4.8-expressing negative control. A potential explanation for this is that short oligosaccharides released by *Tm*cel12A enzymes that were secreted by optimal secretion colonies, such as colonies J and V, have diffused within the agar plate and facilitated the growth of suboptimal secretion colonies.

Nevertheless, the secretion capabilities of the colonies were ranked by dividing the radius of the zone of clearing by the radius of the previously measured colonies. The replicate colonies that corresponded to the six highest ranking colonies from the congo red assay were inoculated into 5 ml cultures and plasmid DNA was prepared. Prepared plasmid DNA was sequenced (Eurofins Genomics) at the signal peptide fragment site using the conventional M13(-21)F primer, which anneals upstream of the pUCGXXX expression fragment.

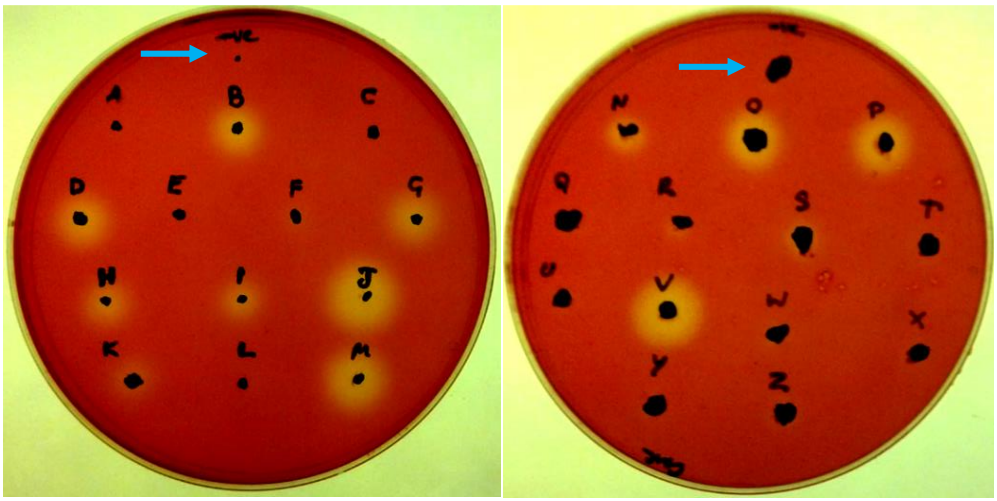


Figure 59 – Congo Red staining for colonies optimal secretion. G. thermoglucosidasius TM242 colonies grown on ASM minimal media agar containing 1% CMC as a sole carbon source, supplemented with 12.5 µg/ml kanamycin. Prior to staining with 1% Congo Red, the colonies were given character identifiers and shaded. 20 minutes staining with Congo red was followed by 20 minutes of destaining with 1M NaCl. Negative controls of G. thermoglucosidasius TM242 expressing signal peptides with no downstream cellulase are presented by blue arrows.

Table 18 – Sequencing results from the six selected colonies from the metachromatic agar assay (Figure 59)

Colony Identifier	Sequenced SigPep	Annotated protein name	Accession
Tmcel12A_colony B_F	SigPep 45	S-Layer Protein	peg3494
Tmcel12A_colony D_F	SigPep 45	S-Layer Protein	peg3494
Tmcel12A_colony J_F	SigPep 18	Subtilisin/Thermolysin	peg3497
Tmcel12A_colony M_F	SigPep 45	S-Layer Protein	peg3494
Tmcel12A_colony O_F	SigPep 08	Oligoendopeptidase F	peg1500
Tmcel12A_colony V_F	SigPep 18	Subtilisin/Thermolysin	peg3497

Two considerations can be taken from the sequencing results of the six colonies selected from the congo red agar assays **(Table 18)**. Firstly, since the metachromatic agar assay of 26 colonies revealed the presence of at least 3 colonies expressing Signal Peptide 45, and 2 colonies expressing Signal Peptide 18, it was hypothesised that the composition of the signal peptide library mixture (pUCG4.8_RPLS_SPL1.0) may be skewed in proportionality in favour of some signal peptides over others. To circumvent this issue, a plasmid construct can be assembled whereby the signal peptide library components are concatemerised into the plasmid assembly to maintain their equimolar concentration.

The second consideration is that the signal peptides that conferred the best (colonies J and V) and second-best secretion rates (colonies B, D and M) were that of the extracellular subtilisin protein (peg3497) and primary S-layer protein (peg3494), respectively. Intriguingly, mass spectrometry analysis of the secretome of *G. thermoglucosidasius* C56-YS93 by another PhD student at the University of Bath, Alexandria Holland, reported that the primary S-layer protein is the most abundant extracellular protein, and the subtilisin protein being the second-most abundant extracellular protein (data not shown). Moreover, the peg3497 and peg3494 ORFs exhibited the highest and third highest consolidated Geosec prediction scores **(Table 17)**. It is enticing to speculate that the absence of "one-size-fits-all" signal peptide may not be a limitation of *G. thermoglucosidasius* TM242. However, this signal peptide library must be improved and exploited for the optimisation of the secretion of many other cellulases before any strong hypotheses can be made.

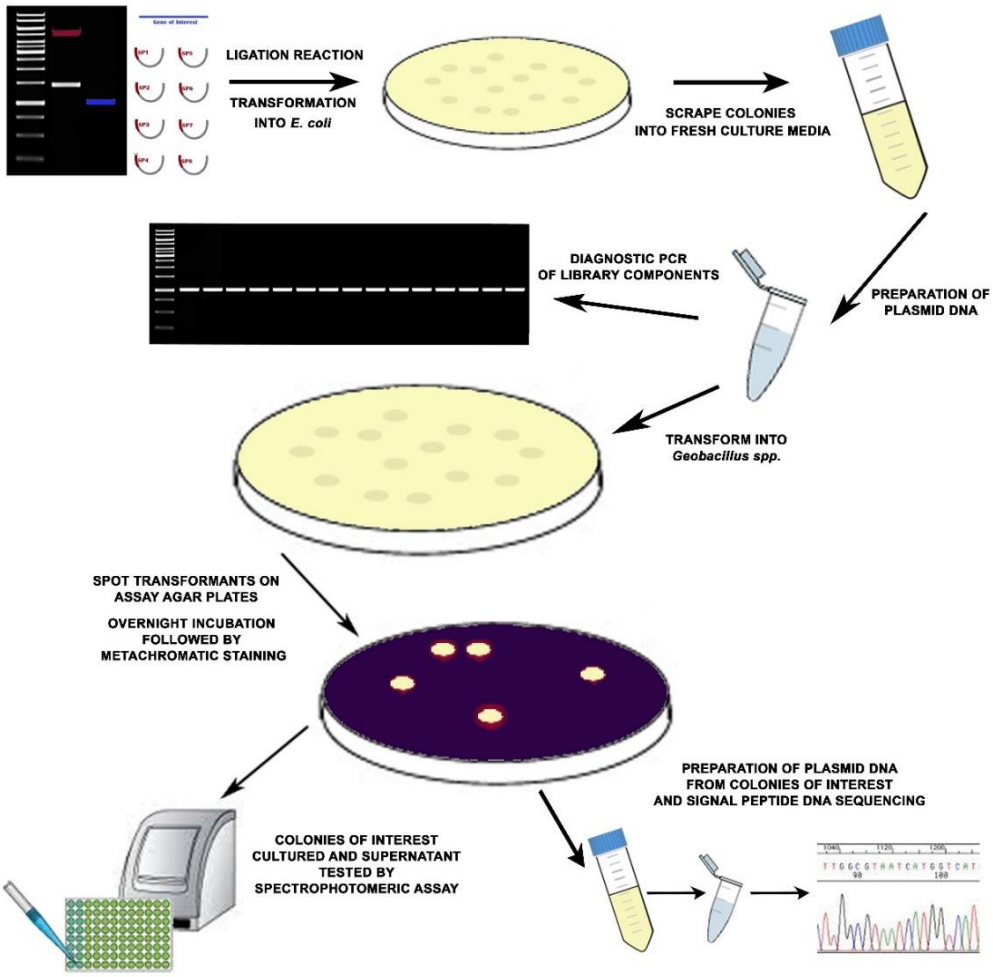


Figure 60 – Schematic of signal peptide library screening for optimal protein secretion.

In conclusion, a *G. thermoglucosidasius* compatible signal peptide library has successfully provided a selection of cellulase-secreting G. thermoglucosidasius TM242 strains with optimal secretion activities. The bioinformatics analyses and the investigation of *Geobacillus* secretion machinery has provided fundamental insights into the similarities of Geobacillus secretion with that of the well-studied *Bacillus subtilis*. However, further studies must be performed with the signal peptide library to elucidate the direct effect that alterations in paired signal peptide can have on extracellular cellulose activities.

Chapter 7: Conclusions and Future Work

The hydrolysis of lignocellulosic feedstocks with commercial enzymes is a major cost in the production of second-generation chemicals/fuels, which significantly affects process economics. Consolidated Bioprocessing (CBP), a potential alternative to conventional processing, could reduce these costs by partially or completely replacing the need for commercial enzymes. This, however, requires the heterologous expression of glycosyl hydrolase in the host strain which, in this study, was *G. thermoglucosidasius* TM242. The work presented in fthis thesis has used *G. thermoglucosidasius* TM242 as a model to investigate the economic feasibility of CBP, constructing a platform for, and evaluating the benefits of heterologous expression and secretion for glycosyl hydrolases from the fermentative thermophile.

Synthetic Biology approaches to CBP engineering.

Since the primary strategy of the CBP project fundamentally depended on the heterologous expression and secretion of glycosyl hydrolases, there was a need to identify and characterise putative promoters, signal peptides and the native utilisation capabilities of *G. thermoglucosidasius* TM242.

Recombinant expression in *G. thermoglucosidasius* TM242

The development of a seamless DNA assembly system, initially using expression reporter constructs was described in Chapter 3. Prior to this project, such constructs were assembled in the pUCGXXX system by conventional restriction digest and ligation of the amplified expression fragment, composed of a promoter and ribosome-binding-site (RBS), upstream of the sfGFP gene. However, to circumvent issues regarding low cloning frequencies, and the limitations of using unique restriction sites (which might be present in the fragment to be cloned), the GoldenGate method of cloning was exploited to construct a variant of pUCG4.8 that facilitates scar-less targeted insertion of expression fragments upstream of the sfGFP gene. This GoldenGate process was used for the successful assembly of six expression constructs.

The first of these promoters was the wild-type version of the *G. thermoglucosidasius* NCIMB 11955 RPLS promoter (pRPLS), which had its native ribosomal binding site (RBS) replaced with an optimal RBS of a *G. stearothermophilus ldh* gene (obtained from Ben Reeve,

Imperial College London). As expected, the RplS promoter exhibited the same strong, constitutive activity under aerobic growth conditions as had been previously characterised in a separate expression system (Reeve, B. personal communication). Superfolder GFP expression levels reached 840 absorbance units (AUs) at a measured culture OD_{600nm} of 1.22, and a linear relationship was observed between OD_{600nm} under aerobic conditions and sfGFP fluorescence, which is consistent with constitutive expression. Strong, constitutive expression made this an obvious candidate promoter for the heterologous expression of glycosyl hydrolases from *G. thermoglucosidasius* TM242. Therefore, the universal vector construct for the expression of glycosyl hydrolases in Chapters 4-6 included expression from the RPLS promoter. Moreover, the RPLS promoter has been made available for, and is currently in use, in a number of different projects, including the expression of downstream components of a terpene-production pathway in *G. thermoglucosidasius*.

The second promoter characterised was a mutant of the uracil phosphoribosyl-transferase (UPRT) promoter (pUP2n38), which had previously been demonstrated to be constitutive, but with weak-to-moderate expression (also from Ben Reeve, Imperial College London). This weak activity was indeed confirmed, with sfGFP fluorescence of only 69 ± 2 AU for a culture OD_{600nm} of 1.16, which was equivalent to 8.2% of the expression conferred by the pRPLS_sfGFP constructs at the same culture OD_{600nm}. Although the promoter was constitutive, its weak activity was deemed insufficient for effective glycosyl hydrolase expression. However, it was utilised for the expression of the synergistic expansin-like protein *Cc*YoaJ, and the cell surface attached cellulase *Cs*cel5SLH (Chapters 4-5).

Sugar utilisation in *G. thermoglucosidasius* TM242

G. thermoglucosidasius TM242 can utilise the major monosaccharides and disaccharides released from complete hydrolysis of *Miscanthus* x *giganteus* (D-glucose, D-xylose, L-arabinose and cellobiose). Analysis of the hemicellulose utilisation island on the *G. thermoglucosidasius* TM242 chromosome revealed intriguing differences between the hemicellulose utilisation potential of *G. thermoglucosidasius* TM242 and *G. thermoglucosidasius* C56-YS93. In contrast to C56-YS93, TM242 encodes an intracellular arabinofuranosidase, associated oligomer transport functions and metabolism of L-arabinose, although the strain does not appear to have the gene cluster encoding extracellular arabinan degradation enzymes (arabinases). The knock-in introduction of this seven-gene mini-operon in *G. thermoglucosidasius* TM242 may therefore be effective in providing significant arabinan-utilisation capabilities.

However, the TM242 strain lacks the gene encoding the extracellular xylanase, *Gt*XynA which is present in C56-YS93, and therefore required the heterologous expression of XynA to facilitate growth on xylan polysaccharides. This has been investigated and achieved by a former PhD student at the University of Bath, Dr. Giannina Espina Silva. The *Gt*XynA-secreting strain demonstrated impressive xylanolytic activity and is currently the main xylan-degrading strain in the current *G. thermoglucosidasius* CBP arsenal of strains.

No genes were discovered in the *G. thermoglucosidasius* TM242 genome that encode extracellular cellulases. Thus, one of the challenges for this project was the selection and heterologous expression of competent secreted exoglucanases and endoglucanases that would confer cellulolytic capabilities on *G. thermoglucosidasius* TM242. To facilitate the construction of plasmid constructs for glycosyl hydrolase expression and secretion, a similar GoldenGate-based plasmid assembly system was designed for the introduction of mature protein coding sequence (MCS) fragments downstream of the universal RPLS promoter and *G. thermoglucosidasius* C56-YS93 xylanase signal peptide.

Exoglucanases

The correct assembly of exoglucanase-expressing constructs was more difficult than that for endoglucanase expression. This phenomenon has previously been reported in studies of *T. fusca* exoglucanases (Irwin *et al.*, 2000), but may simply be due to the larger size of exoglucanase genes compared to those of endoglucanases. This was particularly problematic for this project because exoglucanases are fundamental to the degradation of crystalline cellulose, which is one of the major problems in utilisation of lignocellulosic feedstocks. Nevertheless, three exoglucanase-expressing constructs were eventually assembled into the pUCGXXX; two of which have been characterised in this thesis.

The first exoglucanase, *Ct*cel48S (Cthe_2089), which depolymerises cellulose at its reducing-end, is the most abundant subunit of the *C. thermocellum* cellulosome (Kruus *et al.*, 1995a). However, due to the absence of any CBM domains for exoglucanase attachment to cellulosic substrates, it is not surprising that recombinant *Ct*cel48S was reported to exhibit extremely low specific activity against crystalline Avicel, although it did exhibit higher activity on amorphous phosphoric-acid swollen-cellulose (PASC) (Kruus *et al.*, 1995b). However, this PASC-degrading activity was not detected in supernatants of *Ct*cel48S-secreting *G. thermoglucosidasius* TM242 cultures. Nevertheless, the addition of the non-catalytic *Cc*YoaJ,

which synergistically increases the hydrolytic activity of cellulases, resulted in the highest level of PASC-degrading activity (140.7 U/ml) recorded, indicating that this expansin-like protein may facilitate a similar synergistic function to that of the carbohydrate binding module (CBM3) of the *C. thermocellum* CipA scaffoldin, which improves *Ct*cel48S activity under native conditions.

In contrast, *Tf*cel6B, the non-reducing end exoglucanase from *Thermobifida fusca* XY, does contain a carbohydrate binding module (CBM2), but similarly exhibited weak activity on all cellulosic substrates (Calza *et al.*, 1985). Its merit, however, is the synergism it has been reported to exhibit with the native reducing-end exoglucanase, *Tf*cel48A, and native endoglucanases. It is unfortunate that *G. thermoglucosidasius* TM242 secreting the *Tf*cel6B exhibited notable stuttering of achievable culture OD_{600nm} in minimal media supplemented with glycerol as a sole carbon source; the highest achievable culture OD_{600nm} was 0.58, before reducing thereafter. These signs of metabolic burden are likely to have accrued due to the potential requirement for heterologous *N*-glycosylation of the protein in *G. thermoglucosidasius* TM242; *Tf*cel6B has been characterised to be heavily glycosylated natively (Irwin *et al.*, 2000). Nevertheless, culture supernatants from *Tf*cel6B-secreting *G. thermoglucosidasius* sampled at a culture OD_{600nm} at 0.47 ± 0.03, exhibited impressive PASC-degrading specific activities of 147.6 U/ml, before reducing thereafter. This reduction in extracellular hydrolytic activity was unique to *Tf*cel6B-secreting cultures, but may still be an indication of extracellular proteolytic digestion of extracellular proteins in *G. thermoglucosidasius* TM242.

A number of options may be considered for future engineering of exoglucanase-activities in *G. thermoglucosidasius* TM242. Firstly, a plasmid construct expressing the non-cellulosomal and more active *Ct*cel48Y exoglucanase, which contains a CBM3 domain at its carboxyl-terminal, has been constructed, although its transfer into *G. thermoglucosidasius* TM242 has been unsuccessful to date. The MCS fragment for *Tf*cel48A, the most abundantly expressed protein in *T. fusca* XY, has also been assembled into a holding vector, but transfer into the pUCGXXX expression system has also proven problematic; problems were also encountered when it was originally characterised (Irwin *et al.*, 2000).

The activity of *Ct*cel48S could potentially be improved by substitution of its redundant dockerin domain, which is used for attachment to the native cellulosome, with the CBM3 domain of *Ct*cel48Y. As far as I am aware, the conversion of a cellulosomal-

cellulase into a non-cellulosomal cellulase has not been described, to date. Alternatively, an S-layer protein displaying a corresponding cohesin domain and carbohydrate binding module could be expressed to utilise the present *Ct*cel48S construct.

The exhibited PASC-degrading activity of *Tf*cel6B may be the first empirical indication of glycosylation capabilities in *G. thermoglucosidasius*. However, High Performance Liquid Chromotography (HPLC) and Mass Spectrometry analysis is required to accurately investigate whether secreted *Tf*cel6B is indeed glycosylated by *G. thermoglucosidasius* TM242. The loss of activity of *Tf*cel6B in the culture supernatant needs to be investigated. Potentially it reflects proteolytic activity, which may initially be investigated using protease inhibitors, and subsequently by co-culturing serine-protease inhibitor (Serpin) *Ct*PinA-secreting and *Tf*cel6B-secreting *G. thermoglucosidasius* TM242 strains. *Ct*PinA expression and secretion has been shown to reduce extracellular proteolytic activity of *G. thermoglucosidasius* TM242.

Endoglucanases

Eight endoglucanase genes were successfully assembled into expression constructs, five of which have currently been expressed and characterised in *G. thermoglucosidasius* TM242. This included the three *Thermotoga maritima* endoglucanases *Tm*cel12A, *Tm*cel12B and *Tm*cel5A, the processive *Clostridium thermocellum* endoglucanase *Ct*cel9I, and the cell surface attached *Caldicellulosiruptor saccharolyticus* endoglucanase *Cs*cel5SLH.

Cultures expressing the endoglucanases *Cs*cel5SLH, *Tm*cel12A, *Tm*cel12B and *Tm*cel5A all reached a high OD_{600nm} in minimal media supplemented with glycerol as a sole carbon source, indicating that there was no significant metabolic burden. Effective secretion was also evident from the measured depolymerisation of CMC. Indeed, the culture profiles for *Tm*cel12B- and *Tm*cel5A-secreting strains were virtually identical to that of the negative control strain, with no mature protein coding sequence (MCS). This was not completely unexpected, given the relatively small size of these two *T. maritima* MS8 enzymes (*Tm*cel12B = 282 amino acids, *Tm*cel5A = 348 amino acids). However, although *Tm*cel12A-secreting *G. thermoglucosidasius* cultures exhibited the greatest final culture OD_{600nm} of 5.717 ± 0.17, the maximum growth rate was slower than that of *Tm*cel12B- and *Tm*cel5A-secreting strains. A metabolic burden was observed in the culture growth profile of *G. thermoglucosidasius* secreting *Cs*cel5SLH, the putative cell-surface display endoglucanase, which exhibited slower

growth rates and lower final culture OD_{600nm} than its *T. maritima* endoglucanase-secreting counterparts.

*Tm*cel12B-secreting *G. thermoglucosidasius* had consistently greater CMC-degrading specific activities than its *Tm*cel5A counterpart, consistent with its higher specific activity (890 U/mg) than *Tm*cel5A (616 U/mg). In contrast, *Tm*cel12A-secreting *G. thermoglucosidasius* TM242 exhibited an apparent surge in extracellular CMC-degrading activities between a culture OD_{600nm} of ~1 (192 U/ml) and 5.7 (911 U/ml). Intriguingly, no detectable CMC-degrading activity was observed in supernatants of *Cs*cel5SLH-secreting *G. thermoglucosidasius* collected at a culture OD_{600nm} of 1.0, but supernatant collected at a culture OD_{600nm} of 3.3 ± 0.3 exhibited 632.4 U/ml of CMC-degrading specific activity. The likely explanation is that the S-layer bound endoglucanase was being released as a result of cell lysis. If correct, then this will be the first engineered display of a heterologous protein on the cell surface of *G. thermoglucosidasius*.

The addition of supernatants containing the expansin-like protein *Cc*YoaJ significantly increased the PASC-depolymerising activities of *Tm*cel12A, *Tm*cel12B and *Cs*cel5SLH. However, the study of synergism between cellulases identified that only the combination of Ctcel48S with *Tm*cel12B was significantly synergistic compared to their individual activities. Intriguingly, a potential degree of antagonism was identified between *Tm*cel12A and *Tm*cel12B. Although the additive specific activities of Tmcel12A (67.0 U/ml) and Tmcel12B (44.2 U/ml) should result in a combined specific activity of 111.2 U/ml, the measured specific activity of this *T. maritima* endoglucanase combination was markedly lower at 60.5 U/ml.

Plasmids have been constructed to express the *C. thermocellum* DSM1237 cellulosomal *Ct*cel8A and *Ct*cel9D endoglucanases, which have been shown to exhibit strong activity on amorphous CMC, as well as some weak activity on crystalline substrates (Lee *et al.*, 2011, Schwarz *et al.*, 1986). However, their transfer into *G. thermoglucosidasius* TM242 has proven unsuccessful to date. A plasmid has also been constructed to express the soluble secreted Cscel5 endoglucanase from *Caldicellusiruptor saccharolyticus* DSM8903,which has encountered the same problem. Since *Cs*cel5 is a paralog of *Cs*cel5SLH and is a member of the same glycosyl hydrolase family (GH5), characterisation of a *Cs*cel5-secreting *G. thermoglucosidasius* TM242 may provide some insight into the causes of the metabolic burden of *Cs*cel5SLH-secretion. Confirmation of the binding of *Cs*cel5SLH to the S-layer of *G.*

thermoglucosidasius TM242 needs to be carried out by measurement of endoglucanase activity exhibited by separated soluble and insoluble cell fractions.

The analysis of products released by *Tm*cel12A, which is intracellular and *Tm*cel12B, which is secreted by *T. maritima* MS8, may provide some insight over the roles of these two similar endoglucanases in the utilisation of cellulose by *T. maritima.* This can be done by High-Performance Anion-Exchange Chromatography, which is a powerful tool for the analysis of carbohydrates, including mono-, di-, tri-, oligo- and polysaccharides, sugar alcohols, and amino sugars.

Growth of cellulolytic *G. thermoglucosidasius* on amorphous cellulose

To investigate the growth of cellulolytic *G. thermoglucosidasius* TM242 on amorphous cellulose, co-cultures of endoglucanase-expressing strains were grown in ASM mineral salts medium supplemented with 1.25% CMC as its sole carbon source, in both tube and flask cultures. As expected, the maximum level of growth sustained on CMC was exhibited by the co-culture of *G. thermoglucosidasius* TM242 strains individually expressing *Ct*cel9I, *Tm*cel12A and *Tm*cel12B. The single cultures of *Ct*cel9I-secreting *G. thermoglucosidasius* TM242 reached similar OD_{600nm} values, but the density of the culture could not be sustained. This highlighted *Ct*cel9I as a potential major player in the depolymerisation of amorphous cellulose, which is likely due to its processive activity.

Nevertheless, further investigation of the ability of different combinations of cellulase-expressing *G. thermoglucosidasius* co-cultures to grow on CMC and insoluble substrates, such as PASC and Avicel, must be performed. To facilitate the characterisation of growth profiles on insoluble substrates, the Bradford Assay has been exploited to measure protein concentrations in cellulolytic cultures (Chapter 5).

Developments of the *G. thermoglucosidasius* expression toolkit

RNAseq analysis of the *G. thermoglucosidasius* NCIMB 11955 transcriptome, and the identification of similar promoters in *G. thermoglucosidasius* TM242, to those characterised in *B. subtilis*, has led to the identification of four promoters of interest.

The first promoter was found upstream of the ORF, peg4070, which codes for a hypothetical protein with limited sequence identity to any presently sequenced gene.

Peg4070 transcripts accounted for 0.9-2.5% of total RNA and the p4070 promoter conferred stronger sfGFP expression than that conferred by pUP2n38, but weaker expression than that conferred by RPLS, when characterised under aerobic conditions. Interestingly, in contrast to the pRPLS and pUP2n38 promoters, sfGFP expression increased once the culture had entered stationary phase, and disproportional increases in sfGFP fluorescence were observed during growth. This was illustrated by a strong exponential relationship between increases in sfGFP expression and OD_{600nm} during growth. Although one suggestion for the exponential relationship was continued expression of sfGFP in non-dividing cells, this was not observed in non-dividing cells under oxygen-limiting conditions, so further analysis is required to characterise this activity. Nevertheless, the p4070 promoter characteristics differ from those of pRplS which make it a useful addition to the toolkit.

Another promoter of interest was that of the acetaldehyde dehydrogenase gene (pAdhE), which was identified by the RNAseq analysis to exhibit the greatest increase in transcripts during the aerobic-anaerobic transition, going from no expression under aerobic growth conditions to sfGFP fluorescence levels of 69 ± 0 AUs at a culture OD_{600nm} of 1.083. Although induced promoter activity was weak, the low basal activity and strong inducibility of pAdhE may prove useful in expression studies of toxic or oxygen-sensitive proteins. To fully quantify pAdhE expression under oxygen-limiting conditions and under fully anaerobic conditions, an expression reporter that does not require molecular oxygen for maturation, such as the catechol-2,3-dioxygenase-expressing *pheB* gene, is required (Bartosiak-Jentys *et al.*, 2012). A pUCG4.8_pAdhE_*pheB* construct has been constructed, and its characterisation is the natural step forward for complete characterisation of the inducibility of this promoter.

The trehalose-inducible promoter, pTre, was identified by analysis of the *G. thermoglucosidasius* TM242 genome, as part of a similar *trePAR* operon to that characterised in *B. subtilis* 168. Analysis of the region upstream of the first gene in the operon, *treP,* revealed the presence of almost identical inverted repeats to trePO1 and trePO2, the characterised operator sequences of *B. subtilis* 168 TreR repressor protein. However, unlike the weak expression exhibited by the *B. subtilis* 168 pTre promoter, its *G. thermoglucosidasius* TM242 counterpart exhibited strong expression levels, with sfGFP fluorescence of 500 AU at a culture OD_{600nm} of 1. This was 75% of that observed with expression driven by the RPLS promoter (700 AU), which contains an optimised RBS. Expression of pTre is subject to catabolite repression making it a useful reporter to study of this process. However, the addition of 5 mM trehalose was toxic to cell growth, producing a

dramatic drop in culture OD_{600nm} so the optimal trehalose concentration for induction needs to be determined.

Perhaps the most useful inducible promoter characterised in this project was the sorbitol-inducible promoter, pSorb. Although this seems to control a galactitol utilisation operon in *G. thermoglucosidasius* TM242, the organism did not grow on galactitol. However, it used its stereoisomer, sorbitol (glucitol). Low levels of sfGFP fluorescence (< 10 AUs) were observed in the absence of sorbitol under aerobic growth conditions but the addition of 5 mM sorbitol resulted in sfGFP expression at 80 AU. Addition of 5 mM galactitol did not increase sfGFP expression. Further characterisation of the inducibility of the pSorb promoter is required with different sugar alcohols to uncover its mechanism. In addition, the characterisation of pSorb expression with the addition of keto-D-fructose or β-D-fructofuranose, downstream intermediates of sorbitol metabolism, may provide useful insights. Nevertheless, the strong inducibility of the promoter by sorbitol addition makes it a powerful tool for transient expression of heterologous proteins in *Geobacillus* spp.

Final Conclusions

This project has demonstrated the first example of the genetic engineering of a cellulolytic *Geobacillus* spp. and the first example of a synergistic expansin-like protein functional at elevated temperatures (>55°C). Using these to investigate the first three stages of CBP engineering provides an evaluation of the difficulties of applying CBP as a bioprocess strategy in *G. thermoglucosidasius*. Whereas the GoldenGate-based cloning system facilitated the assembly of 6 expression reporter constructs and 11 cellulase-expression constructs, the process of concept-to-expression in the genetic engineering of *G. thermoglucosidasius* TM242 remains rate-limiting due to problems with expression of the heterologous proteins or, in some cases, obtaining stable constructs in *Geobacillus* sp. Expression of endoglucanases from the strongest characterised promoter was apparently insufficient to provide strong growth on amorphous cellulose, although there is room for improvement with supply on ancilliary proteins such as expansins. Therefore, strong and sustained growth on pure crystalline cellulose remains an ambitious target, given the problems with exoglucanase expression and low activity. Here, a different strategy may be useful, such as the use of lytic polysaccharide monooxygenases to generate a larger number of ends for combined attack by exo and endo-enzymes

This project has also provided a number of tools for heterologous protein secretion in *Geobacillus* spp., including promoters induced in the presence of trehalose or sorbitol, characterisation of the predicted secretomes of *Geobacillus* spp., and a signal peptide library for the screening of the optimal components for protein expression and secretion. Even if complete CBP is not feasible due to the recalcitrance of crystalline cellulose, the provision of additional capacity to degrade solubilised oligomers will add to the catabolic versatility of these strains and can potentially reduce, if not eliminate, the cost of commercially supplied enzymes.

Bibliography

Adams, D. and Ribbons, D. W. (1988) 'The metabolism of aromatic compounds by thermophilicBacilli', *Applied Biochemistry and Biotechnology,* 17(1-3), pp. 231-244.

Aldridge, C., Spence, E., Kirkilionis, M. A., Frigerio, L. and Robinson, C. (2008) 'Tat-dependent targeting of Rieske iron-sulphur proteins to both the plasma and thylakoid membranes in the cyanobacterium Synechocystis PCC6803', *Mol. Microbiol.,* 70, pp. 140-150.

Alper, H., Moxley, J., Nevoigt, E., Fink, G. R. and Stephanopoulos, G. (2006) 'Engineering yeast transcription machinery for improved ethanol tolerance and production', *Science,* 314(5805), pp. 1565-1568.

Alvira, P., Tomas-Pejo, E., Ballesteros, M. and Negro, M. J. (2010) 'Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review', *Bioresource Technology,* 101(13), pp. 4851-4861.

Anderson, J., Dueber, J. E., Leguia, M., Wu, G. C., Goler, J. A., Arkin, A. P. and Keasling, J. D. (2010) 'BglBricks: A flexible standard for biological part assembly', *Journal of biological engineering,* 4(1), pp. 1-12.

Arfi, Y., Shamshoum, M., Rogachev, I., Peleg, Y. and Bayer, E. A. (2014) 'Integration of bacterial lytic polysaccharide monooxygenases into designer cellulosomes promotes enhanced cellulose degradation', *Proceedings of the National Academy of Sciences of the United States of America,* 111(25), pp. 9109-9114.

Arai, T., Araki, R., Tanaka, A., Karita, S., Kimura, T., Sakka, K., & Ohmiya, K. (2003). Characterization of a cellulase containing a family 30 carbohydrate-binding module (CBM) derived from *Clostridium thermocellum* CelJ: importance of the CBM to cellulose hydrolysis. Journal of bacteriology, 185(2), 504-512.

Banerjee, G., Car, S., Scott-Craig, J. S., Borrusch, M. S. and Walton, J. D. (2010) 'Rapid optimization of enzyme mixtures for deconstruction of diverse pretreatment/biomass feedstock combinations', *Biotechnology for biofuels,* 3(1), pp. 22.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K. and Miyano, S. (2002) 'Extensive feature detection of N-terminal protein sorting signals', *Bioinformatics,* 18(2), pp. 298-305.

Barnett, J. P., Eijlander, R. T., Kuipers, O. P. and Robinson, C. (2008) 'A minimal Tat system from a Gram-positive organism: a bifunctional TatA subunit participates in discrete TatAC and TatA complexes', *J Biol Chem,* 283(5), pp. 2534-42.

Bartosiak-Jentys, J., Eley, K. and Leak, D. J. (2012) 'Application of pheB as a reporter gene for Geobacillus spp., enabling qualitative colony screening and quantitative analysis of promoter strength', *Appl Environ Microbiol,* 78(16), pp. 5945-7.

Bartosiak-Jentys, J., Hussein, A. H., Lewis, C. J. and Leak, D. J. (2013) 'Modular system for assessment of glycosyl hydrolase secretion in *Geobacillus thermoglucosidasius*', *Microbiology,* 159(Pt 7), pp. 1267-1275.

Bayer, E. A., Lamed, R., White, B. A. and Flint, H. J. (2008) 'From Cellulosomes to Cellulosomics', *Chemical Record,* 8(6), pp. 364-377.

Beguin, P., Cornet, P., & Millet, J. (1983). Identification of the endoglucanase encoded by the celB gene of *Clostridium thermocellum*. Biochimie, 65(8), 495-500.

Bendtsen, J. D., Nielsen, H., von Heijne, G. and Brunak, S. (2004) 'Improved prediction of signal peptides: SignalP 3.0', *Journal of Molecular Biology,* 340(4), pp. 783-795.

Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T. and Brunak, S. (2005) 'Prediction of twin-arginine signal peptides', *Bmc Bioinformatics,* 6.

Berger, E., Zhang, D., Zverlov, V. V. and Schwarz, W. H. (2007) 'Two noncellulosomal cellulases of Clostridium thermocellum, Cel9I and Cel48Y, hydrolyse crystalline cellulose synergistically', *FEMS microbiology letters,* 268(2), pp. 194-201.

Bhat, M. K. and Bhat, S. (1997) 'Cellulose degrading enzymes and their potential industrial applications', *Biotechnology Advances,* 15(3–4), pp. 583-620.

Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., Nitz, M., Rose, D. R., Tull, D., Wakarchuck, W. W., Wang, Q., Warren, R. A. J., White, A. and Withers, S. G. (1998) 'Mechanisms of cellulases and xylanases', *Biochemical Society Transactions,* 26(2), pp. 156-160.

Blanchard, K., Robic, S. and Matsumura, I. (2014) 'Transformable facultative thermophile Geobacillus stearothermophilus NUB3621 as a host strain for metabolic engineering', *Appl Microbiol Biotechnol,* 98(15), pp. 6715-23.

Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Soon Lee, T., Tullman-Ercek, D., Voigt, C. A., Simmons, B. A. and Keasling, J. D. (2011) 'Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli', *Proceedings of the National Academy of Sciences,* 108(50), pp. 19949-19954.

Bolhuis, A., Broekhuizen, C. P., Sorokin, A., van Roosmalen, M. L., Venema, G., Bron, S., Quax, W. J. and van Dijl , J. M. (1998) 'SecDF of Bacillus subtilis, a Molecular Siamese Twin Required for the Efficient Secretion of Proteins', *Journal of Biological Chemistry,* 273(33), pp. 21217-21224.

Brockmeier, U., Caspers, M., Freudl, R., Jockwer, A., Noll, T. and Eggert, T. (2006) 'Systematic Screening of All Signal Peptides from Bacillus subtilis: A Powerful Strategy in Optimizing Heterologous Protein Secretion in Gram-positive Bacteria', *Journal of Molecular Biology,* 362(3), pp. 393-402.

Bronnenmeier, K., Kern, A., Liebl, W. and Staudenbauer, W. L. (1995) 'PURIFICATION OF THERMOTOGA-MARITIMA ENZYMES FOR THE DEGRADATION OF CELLULOSIC MATERIALS', *Applied and Environmental Microbiology,* 61(4), pp. 1399-1407.

Bunnell, K., Rich, A., Luckett, C., Wang, Y.-J., Martin, E. and Carrier, D. J. (2013) 'Plant maturity effects on the physicochemical properties and dilute acid hydrolysis of switchgrass (Panicum virgatum, L.) hemicelluloses', *ACS Sustainable Chemistry & Engineering,* 1(6), pp. 649-654.

Calza, R. E., Irwin, D. C. and Wilson, D. B. (1985) 'Purification and characterization of two .beta.-1,4-endoglucanases from Thermomonospora fusca', *Biochemistry,* 24(26), pp. 7797-7804.

Chandel, A. K. and Singh, O. V. (2011) 'Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'', *Appl Microbiol Biotechnol,* 89(5), pp. 1289-303.

Chen, H. (2014) *Biotechnology of Lignocellulose: Theory and Practice.* Springer Netherlands.

Chen, X.-a., Ishida, N., Todaka, N., Nakamura, R., Maruyama, J.-i., Takahashi, H. and Kitamoto, K. (2010) 'Promotion of efficient Saccharification of crystalline cellulose by Aspergillus fumigatus Swo1', *Appl Environ Microbiol,* 76(8), pp. 2556-61.

Chen, X.-G., Stabnikova, O., Tay, J.-H., Wang, J.-Y. and Tay, S.-L. (2004) 'Thermoactive extracellular proteases of Geobacillus caldoproteolyticus, sp. nov., from sewage sludge', *Extremophiles,* 8(6), pp. 489-498.

Chundawat, S. P. S., Beckham, G. T., Himmel, M. E. and Dale, B. E. (2011) 'Deconstruction of Lignocellulosic Biomass to Fuels and Chemicals', *Annual Review of Chemical and Biomolecular Engineering, Vol 2 Annual Review of Chemical and Biomolecular Engineering*. Palo Alto: Annual Reviews, pp. 121-145.

Chung, C. T., Niemela, S. L. and Miller, R. H. (1989) 'ONE-STEP PREPARATION OF COMPETENT ESCHERICHIA-COLI - TRANSFORMATION AND STORAGE OF BACTERIAL-

CELLS IN THE SAME SOLUTION', *Proceedings of the National Academy of Sciences of the United States of America,* 86(7), pp. 2172-2175.
Chung, D., Cha, M., Guss, A. M. and Westpheling, J. (2014) 'Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii', *Proceedings of the National Academy of Sciences,* 111(24), pp. 8931-8936.
Clifton-Brown, J. C., Breuer, J. and Jones, M. B. (2007) 'Carbon mitigation by the energy crop, Miscanthus', *Global Change Biology,* 13(11), pp. 2296-2307.
Comino, P., Collins, H., Lahnstein, J., Beahan, C. and Gidley, M. J. (2014) 'Characterisation of soluble and insoluble cell wall fractions from rye, wheat and hull-less barley endosperm flours', *Food Hydrocolloids,* 41, pp. 219-226.
Corbion 2014. 2014 Annual Report. Corbion nv, P.O. Box 349 1000, AH Amsterdam, The Netherlands.
Courtin, C. and Delcour, J. A. (2002) 'Arabinoxylans and endoxylanases in wheat flour bread-making', *Journal of cereal science,* 35(3), pp. 225-243.
Cripps, R. E., Eley, K., Leak, D. J., Rudd, B., Taylor, M., Todd, M., Boakes, S., Martin, S. and Atkinson, T. (2009) 'Metabolic engineering of *Geobacillus thermoglucosidasius* for high yield ethanol production', *Metabolic Engineering,* 11(6), pp. 398-408.
da Silveira, R. L. F. and Mattos, F. L. 'Price And Volatility Transmission In Livestock And Grain Markets: Examining The Effect Of Increasing Ethanol Production Across Countries'. *2015 AAEA & WAEA Joint Annual Meeting, July 26-28, San Francisco, California*: Agricultural and Applied Economics Association & Western Agricultural Economics Association.
Davies, G. and Henrissat, B. (1995) 'Structures and mechanisms of glycosyl hydrolases', *Structure,* 3(9), pp. 853-859.
de Keyzer, J., van der Does, C. and Driessen, A. J. (2003) 'The bacterial translocase: a dynamic protein channel complex', *Cell Mol Life Sci,* 60(10), pp. 2034-52.
De Maayer, P., Brumm, P. J., Mead, D. A. and Cowan, D. A. (2014) 'Comparative analysis of the Geobacillus hemicellulose utilization locus reveals a highly variable target for improved hemicellulolysis', *BMC genomics,* 15(1), pp. 836.
Deguchi, S., Tsujii, K. and Horikoshi, K. (2008) 'Crystalline-to-amorphous transformation of cellulose in hot and compressed water and its implications for hydrothermal conversion', *Green Chemistry,* 10(2), pp. 191-196.
del Solar, G., Giraldo, R., Ruiz-Echevarría, M. J., Espinosa, M. and Díaz-Orejas, R. (1998) 'Replication and control of circular bacterial plasmids', *Microbiology and molecular biology reviews,* 62(2), pp. 434-464.
Demain, A. L., Newcomb, M. and Wu, J. H. D. (2005) 'Cellulase, clostridia, and ethanol', *Microbiology and Molecular Biology Reviews,* 69(1), pp. 124-+.
Demirbas, A. (2009) *Biofuels securing the planet's future energy needs*. New York; London: Springer
Deshpande, M. V. (1992) 'ETHANOL-PRODUCTION FROM CELLULOSE BY COUPLED SACCHARIFICATION FERMENTATION USING SACCHAROMYCES-CEREVISIAE AND CELLULASE COMPLEX FROM SCLEROTIUM-ROLFSII UV-8 MUTANT', *Applied Biochemistry and Biotechnology,* 36(3).
Doi, R. H. and Kosugi, A. (2004) 'Cellulosomes: Plant-cell-wall-degrading enzyme complexes', *Nature Reviews Microbiology,* 2(7), pp. 541-551.
Driessen, A. J., Manting, E. H. and van der Does, C. (2001) 'The structural basis of protein targeting and translocation in bacteria', *Nat Struct Biol,* 8(6), pp. 492-8.
Espina, G., Eley, K., Pompidor, G., Schneider, T. R., Crennell, S. J. and Danson, M. J. (2014) 'A novel-xylosidase structure from *Geobacillus thermoglucosidasius*: the first crystal structure of a glycoside hydrolase family GH52 enzyme reveals unpredicted similarity to other glycoside hydrolase folds', *Acta Crystallographica Section D: Biological Crystallography,* 70(5), pp. 1366-1374.

Extance, J., Crennell, S. J., Eley, K., Cripps, R., Hough, D. W. and Danson, M. J. (2013) 'Structure of a bifunctional alcohol dehydrogenase involved in bioethanol generation in *Geobacillus thermoglucosidasius*', *Acta Crystallographica Section D: Biological Crystallography,* 69(10), pp. 2104-2115.

Forsberg, Z., Vaaje-Kolstad, G., Westereng, B., Bunæs, A. C., Stenstrøm, Y., MacKenzie, A., Sørlie, M., Horn, S. J. and Eijsink, V. G. H. (2011) 'Cleavage of cellulose by a CBM33 protein', *Protein Science : A Publication of the Protein Society,* 20(9), pp. 1479-1483.

Forst, S., Kalve, I. and Durski, W. (1995) 'Molecular analysis of OmpR binding sequences involved in the regulation of ompF in Escherichia coli', *FEMS microbiology letters,* 131(2), pp. 147-151.

Gallezot, P. (2012) 'Conversion of biomass to selected chemical products', *Chemical Society Reviews,* 41(4), pp. 1538-1558.

Gilad, R., Rabinovich, L., Yaron, S., Bayer, E. A., Lamed, R., Gilbert, H. J. and Shoham, Y. (2003) 'CelI, a noncellulosomal family 9 enzyme from Clostridium thermocellum, is a processive endoglucanase that degrades crystalline cellulose', *Journal of bacteriology,* 185(2), pp. 391-398.

Gomi, M., Sonoyama, M. and Mitaku, S. (2004) 'High performance system for signal peptide prediction: SOSUIsignal', *Bioinformatics,* 4, pp. 142-147.

Goosens, V. J., De-San-Eustaquio-Campillo, A., Carballido-López, R. and van Dijl, J. M. (2015) 'A Tat ménage à trois - The role of Bacillus subtilis TatAc in twin-arginine protein translocation', *Biochim Biophys Acta,* 1853(10 Pt A), pp. 2745-53.

Graninger, M., & Messner, P. (2001). Prokaryotic glycosylation. Proteomics, 1, 248-261

Güracar, S. (2011) *Production, purification and characterization of thermostable protease from alkaliphilic and thermophilic geobacillus sp.* İzmir Institute of Technology.

Ha, S.-J., Galazka, J. M., Kim, S. R., Choi, J.-H., Yang, X., Seo, J.-H., Glass, N. L., Cate, J. H. and Jin, Y.-S. (2011) 'Engineered Saccharomyces cerevisiae capable of simultaneous cellobiose and xylose fermentation', *Proceedings of the National Academy of Sciences,* 108(2), pp. 504-509.

Hahn-Hagerdal, B., Karhumaa, K., Fonseca, C., Spencer-Martins, I. and Gorwa-Grauslund, M. F. (2007) 'Towards industrial pentose-fermenting yeast strains', *Applied Microbiology and Biotechnology,* 74(5), pp. 937-953.

Haigler, C. H., Betancur, L., Stiff, M. R. and Tuttle, J. R. (2012) 'Cotton fiber: a powerful single-cell model for cell wall and cellulose research', *Frontiers in plant science,* 3.

Hansen, A. C., Zhang, Q. and Lyne, P. W. (2005) 'Ethanol–diesel fuel blends—a review', *Bioresource technology,* 96(3), pp. 277-285.

Harvey, F. 2010. Green Biologics strikes deals in China. Financial Times.

Hawumba, J. F., Theron, J. and Brözel, V. S. (2002) 'Thermophilic Protease-Producing Geobacillus from Buranga Hot Springs in Western Uganda', *Current Microbiology,* 45(2), pp. 144-150.

Heaton, E. A., Dohleman, F. G. and Long, S. P. (2008) 'Meeting US biofuel goals with less land: the potential of Miscanthus', *Global Change Biology,* 14(9), pp. 2000-2014.

Heijne, G. (1990) 'The signal peptide', *The Journal of Membrane Biology,* 115(3), pp. 195-201.

Heinl, S., Spath, K., Egger, E. and Grabherr, R. (2011) 'Sequence analysis and characterization of two cryptic plasmids derived from Lactobacillus buchneri CD034', *Plasmid,* 66(3), pp. 159-168.

Helfert, C., Gotsche, S. and Dahl, M. K. (1995) 'Cleavage of trehalose-phosphate in Bacillus subtilis is catalysed by a Phospho-α-(1–1)-glucosidase encoded by the treA gene', *Molecular microbiology,* 16(1), pp. 111-120.

Helm, D. (2011) 'Peak oil and energy policy-a critique', *Oxford Review of Economic Policy,* 27(1).

Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J. and Knox, J. P. (2010) 'Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects', *Proceedings of the National Academy of Sciences,* 107(34), pp. 15293-15298.

Hirokazu, S. (2012) 'Genetic transformation of Geobacillus kaustophilus HTA426 by conjugative transfer of host-mimicking plasmids', *Journal of microbiology and biotechnology,* 22(9), pp. 1279-1287.

Huang, X. P., & Monk, C. (2004). Purification and characterization of a cellulase (CMCase) from a newly isolated thermophilic aerobic bacterium Caldibacillus cellulovorans gen. nov., sp. nov. World journal of Microbiology and Biotechnology, 20(1), 85-92.

Hussein, A. H., Lisowska, B. K. and Leak, D. J. (2015) 'Chapter One-The Genus Geobacillus and Their Biotechnological Potential', *Advances in applied microbiology,* 92, pp. 1-48.

Inderwildi, O. R. and King, D. A. (2009) 'Quo vadis biofuels?', *Energy & Environmental Science,* 2(4), pp. 343-346.

Ingram, L., Conway, T., Clark, D., Sewell, G. and Preston, J. (1987) 'Genetic engineering of ethanol production in Escherichia coli', *Applied and Environmental Microbiology,* 53(10), pp. 2420-2425.

Inoue, H., Nojima, H. and Okayama, H. (1990) 'HIGH-EFFICIENCY TRANSFORMATION OF ESCHERICHIA-COLI WITH PLASMIDS', *Gene,* 96(1), pp. 23-28.

Iqbal, I., Aftab, M. N., Afzal, M., Ur-Rehman, A., Aftab, S., Zafar, A., Ud-Din, Z., Khuharo, A. R., Iqbal, J. and Ul-Haq, I. (2015) 'Purification and characterization of cloned alkaline protease gene of Geobacillus stearothermophilus', *Journal of Basic Microbiology,* 55(2), pp. 160-171.

Irwin, D. C., Zhang, S. and Wilson, D. B. (2000) 'Cloning, expression and characterization of a family 48 exocellulase, Cel48A, from Thermobifida fusca', *European Journal of Biochemistry,* 267(16), pp. 4988-4997.

Jamal, S., Nurizzo, D., Boraston, A. B. and Davies, G. J. (2004) 'X-ray Crystal Structure of a Non-crystalline Cellulose-specific Carbohydrate-binding Module: CBM28', *Journal of Molecular Biology,* 339(2), pp. 253-258.

Jin, M., Balan, V., Gunawan, C. and Dale, B. E. (2011) 'Consolidated Bioprocessing (CBP) Performance of Clostridium phytofermentans on AFEX-Treated Corn Stover for Ethanol Production', *Biotechnology and Bioengineering,* 108(6).

Jin, M., Gunawan, C., Balan, V., Lau, M. W. and Dale, B. E. (2012) 'Simultaneous saccharification and co-fermentation (SSCF) of AFEX (TM) pretreated corn stover for ethanol production using commercial enzymes and Saccharomyces cerevisiae 424A(LNH-ST)', *Bioresource Technology,* 110.

Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003) 'Prediction of lipoprotein signal peptides in Gram-negative bacteria', *Protein Sci,* 12(8), pp. 1652-62.

Juneja, A., Kumar, D. and Murthy, G. S. (2013) 'Economic feasibility and environmental life cycle assessment of ethanol production from lignocellulosic feedstock in Pacific Northwest US', *Journal of Renewable and Sustainable Energy,* 5(2), pp. 023142.

Kananavičiūtė, R., Butaitė, E. and Čitavičius, D. (2014) 'Characterization of two novel plasmids from Geobacillus sp. 610 and 1121 strains', *Plasmid,* 71(0), pp. 23-31.

Kanehisa, M. and Goto, S. (2000) 'KEGG: Kyoto Encyclopedia of Genes and Genomes', *Nucleic Acids Research,* 28(1).

Kataeva, I., Li, X. L., Chen, H., Choi, S. K., & Ljungdahl, L. G. (1999). Cloning and sequence analysis of a new cellulase gene encoding CelK, a major cellulosome component of *Clostridium thermocellum*: evidence for gene duplication and recombination. Journal of bacteriology, 181(17), 5288-5295.

Kedrov, A., Sustarsic, M., de Keyzer, J., Caumanns, J. J., Wu, Z. C. and Driessen, A. J. (2013) 'Elucidating the native architecture of the YidC: ribosome complex', *Journal of molecular biology,* 425(22), pp. 4112-4124.

Kennett, R. and Sueoka, N. (1971) 'Gene expression during outgrowth of Bacillus subtilis spores: The relationship between gene order on the chromosome and temporal sequence of enzyme synthesis', *Journal of molecular biology,* 60(1), pp. 31-44.

Kibbe, W. A. (2007) 'OligoCalc: an online oligonucleotide properties calculator', *Nucleic Acids Res,* 35(Web Server issue), pp. W43-6.

Kosugi, A., Murashima, K., Tamaru, Y. and Doi, R. H. (2002) 'Cell-surface-anchoring role of N-terminal surface layer homology domains of Clostridium cellulovorans EngE', *J Bacteriol,* 184(4), pp. 884-8.

Kruus, K., Wang, W. K., Ching, J. and Wu, J. (1995a) 'Exoglucanase activities of the recombinant Clostridium thermocellum CelS, a major cellulosome component', *Journal of bacteriology,* 177(6), pp. 1641-1644.

Kruus, K., Wang, W. K., Ching, J. and Wu, J. H. (1995b) 'Exoglucanase activities of the recombinant Clostridium thermocellum CelS, a major cellulosome component', *Journal of Bacteriology,* 177(6), pp. 1641-1644.

Kulkarni, N., Shendye, A. and Rao, M. (1999) 'Molecular and biotechnological aspects of xylanases', *Fems Microbiology Reviews,* 23(4), pp. 411-456.

Kumar, R., Singh, S. and Singh, O. V. (2008) 'Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives', *Journal of Industrial Microbiology & Biotechnology,* 35(5), pp. 377-391.

Kurokawa, J., Hemjinda, E., Arai, T., Kimura, T., Sakka, K., & Ohmiya, K. (2002). *Clostridium thermocellum* cellulase CelT, a family 9 endoglucanase without an Ig-like domain or family 3c carbohydrate-binding module. Applied microbiology and biotechnology, 59(4-5), 455-461.

la Grange, D. C., den Haan, R. and van Zyl, W. H. (2010) 'Engineering cellulolytic ability into bioprocessing organisms', *Appl Microbiol Biotechnol,* 87(4), pp. 1195-208.

Lanxess 2014. 2014 Annual Report.

Lau, M. W. and Dale, B. E. (2009) 'Cellulosic ethanol production from AFEX-treated corn stover using Saccharomyces cerevisiae 424A(LNH-ST)', *Proceedings of the National Academy of Sciences,* 106(5), pp. 1368-1373.

Lee, H. J., Lee, S., Ko, H. J., Kim, K. H., & Choi, I. G. (2010). An expansin-like protein from *Hahella chejuensis* binds cellulose and enhances cellulase activity. Molecules and cells, 29(4), 379-385.

Lee, H. L., Chang, C. K., Teng, K. H. and Liang, P. H. (2011) 'Construction and characterization of different fusion proteins between cellulases and beta-glucosidase to improve glucose production and thermostability', *Bioresource Technology,* 102(4), pp. 3973-3976.

Lee, S. K., Chou, H., Ham, T. S., Lee, T. S. and Keasling, J. D. (2008) 'Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels', *Current Opinion in Biotechnology,* 19(6), pp. 556-563.

Leibovitz, E., Ohayon, H., Gounon, P. and Beguin, P. (1997) 'Characterization and subcellular localization of the Clostridium thermocellum scaffoldin dockerin binding protein SdbA', *Journal of bacteriology,* 179(8), pp. 2519-2523.

Lemaire, M. A. R. C., & Béguin, P. I. E. R. R. E. (1993). Nucleotide sequence of the celG gene of *Clostridium thermocellum* and characterization of its product, endoglucanase CelG. Journal of bacteriology, 175(11), 3353-3360.

Levy, I., Shani, Z. and Shoseyov, O. (2002) 'Modification of polysaccharides and plant cell wall by endo-1,4-beta-glucanase and cellulose-binding domains', *Biomolecular Engineering,* 19(1), pp. 17-30.

Lewandowski, I., Clifton-Brown, J., Scurlock, J. and Huisman, W. (2000) 'Miscanthus: European experience with a novel energy crop', *Biomass and Bioenergy,* 19(4), pp. 209-227.

Li, Z. (2009) *Bio-based composites that mimic the plant cell wall.* Virginia Polytechnic Institute and State University.

Liang, C., Xue, Y., Fioroni, M., Rodríguez-Ropero, F., Zhou, C., Schwaneberg, U., & Ma, Y. (2011). Cloning and characterization of a thermostable and halo-tolerant endoglucanase from *Thermoanaerobacter tengcongensis* MB4. Applied microbiology and biotechnology, 89(2), 315-326.

Liao, H. and Kanikula, A. (1990) 'Increased efficiency of transformation ofBacillus stearothermophilus by a plasmid carrying a thermostable kanamycin resistance marker', *Current Microbiology,* 21(5), pp. 301-306.

Liao, H., McKenzie, T. and Hageman, R. (1986) 'Isolation of a thermostable enzyme variant by cloning and selection in a thermophile', *Proceedings of the National Academy of Sciences,* 83(3), pp. 576-580.

LignoWorks 2015. Lignin. URL: http://www.icfar.ca/lignoworks/content/what-lignin.html [Accessed on 21st March 2016]

Limayem, A. and Ricke, S. C. (2012) 'Lignocellulosic biomass for bioethanol production: current perspectives, potential issues and future prospects', *Progress in Energy and Combustion Science,* 38(4), pp. 449-467.

Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., Papanek, B. A., Riley, L. A., Guss, A. M. and Liao, J. C. (2015) 'Consolidated bioprocessing of cellulose to isobutanol using Clostridium thermocellum', *Metabolic Engineering,* 31, pp. 44-52.

Lin, P. P., Rabe, K. S., Takasumi, J. L., Kadisch, M., Arnold, F. H. and Liao, J. C. (2014) 'Erratum to "Isobutanol production at elevated temperatures in thermophilic *Geobacillus thermoglucosidasius*" [Metab. Eng. 24C (2014) 1–8]', *Metabolic Engineering,* 24(0), pp. 192.

Lynd, L. R., van Zyl, W. H., McBride, J. E. and Laser, M. (2005) 'Consolidated bioprocessing of cellulosic biomass: an update', *Current Opinion in Biotechnology,* 16(5), pp. 577-583.

Lynd, L. R., Weimer, P. J., van Zyl, W. H. and Pretorius, I. S. (2002) 'Microbial cellulose utilization: Fundamentals and biotechnology', *Microbiology and Molecular Biology Reviews,* 66(3), pp. 506-+.

Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., Chapman, J., Chertkov, O., Coutinho, P. M., Cullen, D., Danchin, E. G. J., Grigoriev, I. V., Harris, P., Jackson, M., Kubicek, C. P., Han, C. S., Ho, I., Larrondo, L. F., de Leon, A. L., Magnuson, J. K., Merino, S., Misra, M., Nelson, B., Putnam, N., Robbertse, B., Salamov, A. A., Schmoll, M., Terry, A., Thayer, N., Westerholm-Parvinen, A., Schoch, C. L., Yao, J., Barbote, R., Nelson, M. A., Detter, C., Bruce, D., Kuske, C. R., Xie, G., Richardson, P., Rokhsar, D. S., Lucas, S. M., Rubin, E. M., Dunn-Coleman, N., Ward, M. and Brettin, T. S. (2008) 'Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina)', *Nature Biotechnology,* 26(5), pp. 553-560.

Mazzoli, R., Lamberti, C. and Pessione, E. (2012) 'Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies', *Trends Biotechnol: Vol. 2*. England: 2011 Elsevier Ltd, pp. 111-9.

Mesnage, S., Fontaine, T., Mignot, T., Delepierre, M., Mock, M. and Fouet, A. (2000) 'Bacterial SLH domain proteins are non-covalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation', *Embo Journal,* 19(17), pp. 4473-4484.

Mesnage, S., Tosi-Couture, E., Mock, M. and Fouet, A. (1999) 'The S-layer homology domain as a means for anchoring heterologous proteins on the cell surface of Bacillus anthracis', *Journal of Applied Microbiology,* 87(2), pp. 256-260.

Mingardon, F., Bagert, J. D., Maisonnier, C., Trudeau, D. L. and Arnold, F. H. (2011) 'Comparison of family 9 cellulases from mesophilic and thermophilic bacteria', *Appl Environ Microbiol,* 77(4), pp. 1436-42.

Miyake, R., Shigeri, Y., Tatsu, Y., Yumoto, N., Umekawa, M., Tsujimoto, Y., Matsui, H. and Watanabe, K. (2005) 'Two thimet oligopeptidase-like Pz peptidases produced by a collagen-degrading thermophile, Geobacillus collagenovorans MO-1', *J Bacteriol,* 187(12), pp. 4140-8.

Moller, S., Croning, M. D. R. and Apweiler, R. (2001) 'Evaluation of methods for the prediction of membrane spanning regions', *Bioinformatics,* 17(7), pp. 646-653.

Morag, E., Bayer, E. A., & Lamed, R. (1990). Relationship of cellulosomal and noncellulosomal xylanases of *Clostridium thermocellum* to cellulose-degrading enzymes. Journal of Bacteriology, 172(10), 6098-6105.

Naik, S. N., Goud, V. V., Rout, P. K. and Dalai, A. K. (2010) 'Production of first and second generation biofuels: A comprehensive review', *Renewable & Sustainable Energy Reviews,* 14(2), pp. 578-597.

Nakazawa, H., Okada, K., Onodera, T., Ogasawara, W., Okada, H., & Morikawa, Y. (2009). Directed evolution of endoglucanase III (Cel12A) from *Trichoderma reesei*. Applied microbiology and biotechnology, 83(4), 649-657.

Narumi, I., Nakayama, N., Nakamoto, S., Kimura, T., Yanagisawa, T. and Kihara, H. (1993) 'Construction of a new shuttle vector pSTE33 and its stabilities in Bacillus stearothermophilus, bacillus subtilis, and Escherichia coli', *Biotechnology Letters,* 15(8), pp. 815-820.

Narumi, I., Sawakami, K., Nakamoto, S., Nakayama, N., Yanagisawa, T., Takahashi, N. and Kihara, H. (1992) 'A newly isolatedBacillus stearotheromophilus K1041 and its transformation by electroporation', *Biotechnology Techniques,* 6(1), pp. 83-86.

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., Lysenko, A. M., Petrunyaka, V. V., Osipov, G. A., Belyaev, S. S. and Ivanov, M. V. (2001) 'Taxonomic study of aerobic thermophilic bacilli: descriptions of Geobacillus subterraneus gen. nov., sp nov and Geobacillus uzenensis sp nov from petroleum reservoirs and transfer of Bacillus stearothermophilus Bacillus thermocatenulatus, Bacillus thermoleovorans, Bacillus kaustophilus, Bacillus thermoglucosidasius and Bacillus thermodenitrificans to Geobacillus as the new combinations G-stearothermophilus, G-thermocatenulatus, G-thermoleovorans, G-kaustophilus, G-thermoglucosidasius and G-thermodenitrificans', *International Journal of Systematic and Evolutionary Microbiology,* 51, pp. 433-446.

Ng, I. S., Li, C. W., Yeh, Y. F., Chen, P. T., Chir, J. L., Ma, C. H., ... & Tong, C. G. (2009). A novel endo-glucanase from the thermophilic bacterium *Geobacillus* sp. 70PC53 with high activity and stability over a broad range of temperatures. Extremophiles, 13(3), 425-435.

Norrander, J., Kempe, T. and Messing, J. (1983) 'Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis', *Gene,* 26(1), pp. 101-6.

Nouwen, N. and Driessen, A. J. (2002) 'SecDFyajC forms a heterotetrameric complex with YidC', *Molecular microbiology,* 44(5), pp. 1397-1405.

Ou, M. S., Mohammed, N., Ingram, L. O. and Shanmugam, K. T. (2009) 'Thermophilic Bacillus coagulans Requires Less Cellulases for Simultaneous Saccharification and Fermentation of Cellulose to Products than Mesophilic Microbial Biocatalysts', *Applied Biochemistry and Biotechnology,* 155(1-3).

Ozdemir, I., Blumer-Schuette, S. E. and Kelly, R. M. (2012) 'S-Layer Homology Domain Proteins Csac_0678 and Csac_2722 Are Implicated in Plant Polysaccharide

Deconstruction by the Extremely Thermophilic Bacterium Caldicellulosiruptor saccharolyticus', *Applied and Environmental Microbiology,* 78(3), pp. 768-777.
Palmer, T., Sargent, F. and Berks, B. C. (2005) 'Export of complex cofactor-containing proteins by the bacterial Tat pathway', *Trends in microbiology,* 13(4), pp. 175-180.
Papanikou, E., Karamanou, S. and Economou, A. (2007) 'Bacterial protein secretion through the translocase nanomachine', *Nat Rev Micro,* 5(11), pp. 839-851.
Patel, M. A., Ou, M. S., Ingram, L. O. and Shanmugam, K. T. (2005) 'Simultaneous saccharification and co-fermentation of crystalline cellulose and sugar cane bagasse hemicellulose hydrolysate to lactate by a thermotolerant acidophilic Bacillus sp', *Biotechnology Progress,* 21(5).
Pereira, J. H., Chen, Z., McAndrew, R. P., Sapra, R., Chhabra, S. R., Sale, K. L., Simmons, B. A. and Adams, P. D. (2010) 'Biochemical characterization and crystal structure of endoglucanase Cel5A from the hyperthermophilic Thermotoga maritima', *Journal of Structural Biology,* 172(3), pp. 372-379.
Pereira, S. C., Maehara, L., Machado, C. M. M. and Farinas, C. S. (2015) '2G ethanol from the whole sugarcane lignocellulosic biomass', *Biotechnology for Biofuels,* 8, pp. 44.
Pinto, P., Evtuguin, D. and Neto, C. P. (2005) 'Structure of hardwood glucuronoxylans: modifications and impact on pulp retention during wood kraft pulping', *Carbohydrate Polymers,* 60(4), pp. 489-497.
Pohl, M., Mesch, K., Rodenbrock, A. and Kula, M. (1995) 'Stability investigations on on the pyruvate decarboxylases from Zymomonas mobilis', *Biotechnology and applied biochemistry,* 22(1), pp. 95-105.
Pédelacq, J.-D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. (2006) 'Engineering and characterization of a superfolder green fluorescent protein', *Nature biotechnology,* 24(1), pp. 79-88.
Quiroz-Castañeda, R. E., Martínez-Anaya, C., Cuervo-Soto, L. I., Segovia, L. and Folch-Mallol, J. L. (2011) 'Loosenin, a novel protein with cellulose-disrupting activity from Bjerkandera adusta', *Microb Cell Fact,* 10(8).
Rahman, O., Cummings, S., Harrington, D. and Sutcliffe, I. (2008) 'Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria', *World Journal of Microbiology and Biotechnology,* 24(11), pp. 2377-2382.
Rass-Hansen, J., Falsig, H., Jørgensen, B. and Christensen, C. H. (2007) 'Bioethanol: fuel or feedstock?', *Journal of Chemical Technology & Biotechnology,* 82(4), pp. 329-333.
Reverbel-Leroy, C., Pages, S., Belaich, A., Belaich, J. P., & Tardif, C. (1997). The processive endocellulase CelF, a major component of the *Clostridium cellulolyticum* cellulosome: purification and characterization of the recombinant form. Journal of bacteriology, 179(1), 46-52.
(RFA), R. F. A. 2015. GOING GLOBAL: 2015 Ethanol Industry Outlook. URL: http://www.ethanolrfa.org/resources/publications/outlook/ [Accessed on 21st March 2016]
Romero, S., Merino, E., Bolívar, F., Gosset, G. and Martinez, A. (2007) 'Metabolic engineering of Bacillus subtilis for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism', *Applied and environmental microbiology,* 73(16), pp. 5190-5198.
Ryu, S. and Karim, M. (2011) 'A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates', *Applied Microbiology and Biotechnology,* 91(3), pp. 529-542.
Saloheimo, M., Paloheimo, M., Hakola, S., Pere, J., Swanson, B., Nyyssönen, E., Bhatia, A., Ward, M. and Penttilä, M. (2002) 'Swollenin, a Trichoderma reesei protein with sequence similarity to the plant expansins, exhibits disruption activity on cellulosic materials', *Eur J Biochem,* 269(17), pp. 4202-11.

Sampedro, J. and Cosgrove, D. (2005) 'The expansin superfamily', *Genome Biology,* 6(12), pp. 242.

Sandgren, M., Wu, M., Karkehabadi, S., Mitchinson, C., Kelemen, B. R., Larenas, E. A., Ståhlberg, J. and Hansson, H. (2013) 'The Structure of a Bacterial Cellobiohydrolase: The Catalytic Core of the Thermobifida fusca Family GH6 Cellobiohydrolase Cel6B', *Journal of Molecular Biology,* 425(3), pp. 622-635.

Sauve, V., Bruno, S., Berks, B. C. and Hemmings, A. M. (2007) 'The SoxYZ complex carries sulfur cycle intermediates on a peptide swinging arm', *J. Biol. Chem.,* 282, pp. 23194-23204.

Schneewind, O. and Missiakas, D. M. (2012) 'Protein secretion and surface display in Gram-positive bacteria', *Philosophical Transactions of the Royal Society B-Biological Sciences,* 367(1592).

Schuster, B. and Sleytr, U. B. (2014) 'Biomimetic interfaces based on S-layer proteins, lipid membranes and functional biomolecules', *Journal of the Royal Society Interface,* 11(96), pp. 20140232.

Schwarz, W. H. (2001) 'The cellulosome and cellulose degradation by anaerobic bacteria', *Applied Microbiology and Biotechnology,* 56(5-6), pp. 634-649.

Schwarz, W. H., Gräbnitz, F. and Staudenbauer, W. L. (1986) 'Properties of a Clostridium thermocellum Endoglucanase Produced in Escherichia coli', *Applied and Environmental Microbiology,* 51(6), pp. 1293-1299.

Schwarz, W. H. and Zverlov, V. V. (2006) 'Protease inhibitors in bacteria: an emerging concept for the regulation of bacterial protein complexes?', *Molecular microbiology,* 60(6), pp. 1323-1326.

Schöck, F. and Dahl, M. K. (1996) 'Expression of the tre operon of Bacillus subtilis 168 is regulated by the repressor TreR', *Journal of bacteriology,* 178(15), pp. 4576-4581.

Serek, J., Bauer-Manz, G., Struhalla, G., van den Berg, L., Kiefer, D., Dalbey, R. and Kuhn, A. (2004) 'Escherichia coli YidC is a membrane insertase for Sec-independent proteins', *The EMBO journal,* 23(2), pp. 294-301.

Shyamala, S., Ravikumar, S., Vikramathithan, J., & Srikumar, K. (2011). Isolation, purification, and characterization of two thermostable endo-1, 4-β-D-glucanase forms from *Opuntia vulgaris*. Applied biochemistry and biotechnology, 165(7-8), 1597-1610.

Sirko, A., Zehelein, E., Freundlich, M. and Sawers, G. (1993) 'Integration host factor is required for anaerobic pyruvate induction of pfl operon expression in Escherichia coli', *Journal of bacteriology,* 175(18), pp. 5769-5777.

Slade, R. 2010. Low cost enzymes could catalyse a revolution in ethanol production. Centre for Energy Policy and Technology, Imperial College London: Cleantech.

Solovyev, V. and Salamov, A. (2011) 'Automatic annotation of microbial genomes and metagenomic sequences', *Metagenomics and its applications in agriculture, biomedicine and environmental studies*, pp. 61-78.

Spiridonov, N. A. and Wilson, D. B. (1998) 'Regulation of Biosynthesis of Individual Cellulases in Thermomonospora fusca', *Journal of Bacteriology,* 180(14), pp. 3529-3532.

Stephenson, K. (2005) 'Sec-dependent protein translocation across biological membranes: evolutionary conservation of an essential protein transport pathway (Review)', *Molecular membrane biology,* 22(1-2), pp. 17-28.

Suzuki, H., Murakami, A. and Yoshida, K.-i. (2012) 'Counterselection System for Geobacillus kaustophilus HTA426 through Disruption of pyrF and pyrR', *Applied and Environmental Microbiology,* 78(20), pp. 7376-7383.

Suzuki, H., Wada, K., Furukawa, M., Doi, K. and Ohshima, T. (2013a) 'A ternary conjugation system for the construction of DNA libraries for Geobacillus kaustophilus HTA426', *Biosci Biotechnol Biochem,* 77(11), pp. 2316-8.

Suzuki, H., Yoshida, K.-i. and Ohshima, T. (2013b) 'Polysaccharide-Degrading Thermophiles Generated by Heterologous Gene Expression in Geobacillus kaustophilus HTA426', *Applied and Environmental Microbiology,* 79(17), pp. 5151-5158.

Talebnia, F., Karakashev, D. and Angelidaki, I. (2010) 'Production of bioethanol from wheat straw: An overview on pretreatment, hydrolysis and fermentation', *Bioresource Technology,* 101(13), pp. 4744-4753.

Taylor, M. P., Esteban, C. D. and Leak, D. J. (2008) 'Development of a versatile shuttle vector for gene expression in Geobacillus spp', *Plasmid,* 60(1), pp. 45-52.

Telke, A. A., Ghatge, S. S., Kang, S. H., Thangapandian, S., Lee, K. W., Shin, H. D., ... & Kim, S. W. (2012). Construction and characterization of chimeric cellulases with enhanced catalytic activity towards insoluble cellulosic substrates. Bioresource technology, 112, 10-17.

Thompson, A. H., Studholme, D. J., Green, E. M. and Leak, D. J. (2008) 'Heterologous expression of pyruvate decarboxylase in *Geobacillus thermoglucosidasius*', *Biotechnology Letters,* 30(8), pp. 1359-1365.

Tsukimoto, K., Takada, R., Araki, Y., Suzuki, K., Karita, S., Wakagi, T., Shoun, H., Watanabe, T. and Fushinobu, S. (2010) 'Recognition of cellooligosaccharides by a family 28 carbohydrate-binding module', *FEBS letters,* 584(6), pp. 1205-1211.

Tsutsui, Y., Liu, L., Gershenson, A. and Wintrode, P. L. (2006) 'The conformational dynamics of a metastable serpin studied by hydrogen exchange and mass spectrometry', *Biochemistry,* 45(21), pp. 6561-6569.

Vasan, P. T., Piriya, P. S., Prabhu, D. I. G. and Vennison, S. J. (2011) 'Cellulosic ethanol production by Zymomonas mobilis harboring an endoglucanase gene from Enterobacter cloacae', *Bioresour Technol,* 102(3), pp. 2585-9.

Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J. and Siika-Aho, M. (2007) 'Thermostable enzymes in lignocellulose hydrolysis', *Adv Biochem Eng Biotechnol,* 108, pp. 121-45.

Von Heijne, G. (1998) 'Protein transport: Life and death of a signal peptide', *Nature,* 396(6707), pp. 111-113.

von Ossowski, I., Eaton, J. T., Czjzek, M., Perkins, S. J., Frandsen, T. P., Schülein, M., Panine, P., Henrissat, B. and Receveur-Bréchot, V. (2005) 'Protein Disorder: Conformational Distribution of the Flexible Linker in a Chimeric Double Cellulase', *Biophysical Journal,* 88(4), pp. 2823-2832.

Wang, W., Kruus, K. and Wu, J. (1994) 'Cloning and expression of theClostridium thermocellum celS gene inEscherichia coli', *Applied microbiology and biotechnology,* 42(2-3), pp. 346-352.

Wang, Z., Gerstein, M. and Snyder, M. (2009) 'RNA-Seq: a revolutionary tool for transcriptomics', *Nature Reviews Genetics,* 10(1), pp. 57-63.

Wei, H., Fu, Y., Magnusson, L., Baker, J. O., Maness, P.-C., Xu, Q., Yang, S., Bowersox, A., Bogorad, I. and Wang, W. (2014) 'Comparison of transcriptional profiles of Clostridium thermocellum grown on cellobiose and pretreated yellow poplar using RNA-Seq', *Frontiers in microbiology,* 5.

Westers, L., Westers, H. and Quax, W. J. (2004) 'Bacillus subtilis as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism', *Biochimica Et Biophysica Acta-Molecular Cell Research,* 1694(1-3), pp. 299-310.

Withers, S. G. (2001) 'Mechanisms of glycosyl transferases and hydrolases', *Carbohydrate Polymers,* 44(4), pp. 325-337.

Wood, B. E. and Ingram, L. O. (1992) 'ETHANOL-PRODUCTION FROM CELLOBIOSE, AMORPHOUS CELLULOSE, AND CRYSTALLINE CELLULOSE BY RECOMBINANT KLEBSIELLA-OXYTOCA CONTAINING CHROMOSOMALLY INTEGRATED ZYMOMONAS-MOBILIS GENES FOR ETHANOL-PRODUCTION AND PLASMIDS EXPRESSING

THERMOSTABLE CELLULASE GENES FROM CLOSTRIDIUM-THERMOCELLUM', *Applied and Environmental Microbiology,* 58(7).

Wu, T.-H., Huang, C.-H., Ko, T.-P., Lai, H.-L., Ma, Y., Chen, C.-C., Cheng, Y.-S., Liu, J.-R. and Guo, R.-T. (2011) 'Diverse substrate recognition mechanism revealed by Thermotoga maritima Cel5A structures in complex with cellotetraose, cellobiose and mannotriose', *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1814(12), pp. 1832-1840.

Yamane, K., Bunai, K. and Kakeshita, H. (2004) 'Protein traffic for secretion and related machinery of Bacillus subtilis', *Bioscience, biotechnology, and biochemistry,* 68(10), pp. 2007-2023.

Yanase, H., Yamamoto, K., Sato, D. and Okamoto, K. (2005) 'Ethanol production from cellobiose by Zymobacter palmae carrying the Ruminocuccus albus beta-glucosidase gene', *J Biotechnol,* 118(1), pp. 35-43.

Zeigler, D. R. 2001. *The Genus Geobacillus.* Seventh ed. *Bacillus* Genetic Stock Center Catalog of Strains.

Zhang, J., Shao, X., Townsend, O. V. and Lynd, L. R. (2009) 'Simultaneous Saccharification and Co-Fermentation of Paper Sludge to Ethanol by Saccharomyces cerevisiae RWB222-Part I: Kinetic Modeling and Parameters', *Biotechnology and Bioengineering,* 104(5).

Zhang, S., Lao, G. and Wilson, D. B. (1995) 'Characterization of a Thermomonospora fusca exocellulase', *Biochemistry,* 34(10), pp. 3386-95.

Zhang, X. Z., Sathitsuksanoh, N., Zhu, Z. G. and Zhang, Y. H. P. (2011) 'One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic Bacillus subtilis', *Metabolic Engineering,* 13(4), pp. 364-372.

Zhang, Y.-H. P. and Lynd, L. R. (2005) 'Determination of the number-average degree of polymerization of cellodextrins and cellulose with application to enzymatic hydrolysis', *Biomacromolecules,* 6(3), pp. 1510-1515.

Zhu, W., Cha, D., Cheng, G., Peng, Q. and Shen, P. (2007) 'Purification and characterization of a thermostable protease from a newly isolated *Geobacillus* sp. YMTC 1049', *Enzyme and Microbial Technology,* 40(6), pp. 1592-1597.

Zverlov, V. V., Velikodvorskaya, G. V., Schwarz, W. H., Bronnenmeier, K., Kellermann, J., & Staudenbauer, W. L. (1998). Multidomain structure and cellulosomal localization of the *Clostridium thermocellum* cellobiohydrolase CbhA. Journal of bacteriology, 180(12), 3091-3099.

Zverlov, V. V., Velikodvorskaya, G. A., & Schwarz, W. H. (2003). Two new cellulosome components encoded downstream of celI in the genome of *Clostridium thermocellum*: the non-processive endoglucanase CelN and the possibly structural protein CseP. Microbiology, 149(2), 515-524.

Zverlov, V. V., Schantz, N., & Schwarz, W. H. (2005). A major new component in the cellulosome of *Clostridium thermocellum* is a processive endo-β-1, 4-glucanase producing cellotetraose. FEMS microbiology letters, 249(2), 353-358.

APPENDIX I

Table 19 - Primers used for PCR reactions performed in this study. Type IIS restriction sites (either *Bsa*I (GGTCTC) or *BsmBI* (CGTCTC)) are highlighted in bold. pUCGXXX-based restriction sites (*Sal*I, *Cla*I, *Ngo*MIV and *Sac*I) are underlined.

	Primer Name	Primer Sequence
1	*GG_Cla*I_sfGFP_F	ttcccg**GGTCTC**ATCGATgcgtaaaggcgaag
2	*Sac*I_sfGFP_R	ttttGAGCTCgcaaaaaaacgccctttcggggcgcg atcatttgtacagttcatcc
3	*GG_BsaIKO*_AmpR_R	tttt**GGTCTC**ggctcccgcggtatcattgcagcactg g
4	*GG_BsaIKO*_AmpR_F	tttt**GGTCTC**ggagccacgctcaccggctccag
5	*GG_5'prom*_R	aacccg**GGTCTC**tcgacGTCGACtgcaggcatgcaag cttggc
6	*GG_5'Prom*_pRPLS_F	ttt**GGTCTC**GTCGACaatcgttaaagcggacg
7	*GG_Cla*I_pRPLS_R	tttc**GGTCTC**ATCGATtcgaatcactccttatctaga c
8	*GG_5'Prom*_pUP2n38_F	tttt**GGTCTC**gGTCGACgcgtgttttttgttgggag
9	*GG_Cla*I_pUP2n38_R	tttt**GGTCTC**ATCGATtttgcactctcctctttgtt
10	*GG_5'Prom*_p4070_F	tttt**GGTCTC**gGTCGACtgcaagtctctccattgtcc
11	*GG_Cla*I_p4070_R	tttt**GGTCTC**ATCGATatttcctcctcattcattatc
12	*GG_5'Prom*_pAdhE_F	tttt**GGTCTC**gGTCGACgagccggtggcttgaagggc
13	*GG_Cla*I_pAdhE_R	tttt**GGTCTC**ATCGATattccccctttgtatttttg
14	*GG_5'Prom*_pTre_F	ttttGTCGACgtataggatttcatcgggc
15	*GG_Cla*I_pTre_R	ttttATCGATacgaaaacccccctaactcc

16	*GG_5'Prom*_pSorb_F	tttt**GGTCTC**ggtcgaatacaccgccccttgatcgc
17	*GG_Cla*I_pSorb_R	tttt**GGTCTC**ATCGATtcctcctttctctacgcttat tgtagcgc
18	*GG_Pdi*I_*Tf*cel48A_1F	tt**GGTCTC**GCCGGCgcggtggcgtgctcggtgg
19	*GG_Tf*cel48A_1R	tt**GGTCTC**gccgaatcatgtcccggcgcgc
20	*GG_Tf*cel48A_2F	tt**GGTCTC**ttcggcggaagtgaccgtgc
21	*GG_Tf*cel48A_2R	tt**GGTCTC**cacgcccacgcactgctcg
22	*GG_Tf*cel48A_3F	tt**GGTCTC**ggcgtggcgaattggaagcag
23	*GG_Stop_Tf*cel48A_3R	tt**GGTCTC**gcgttaacccgccccaaataac
24	*GG_Pdi*I_*Tfcel6B*_F	tt**GGTCTC**tattgatgcggcgcatcatg
25	*GG_Stop_Tfcel6B*_R	tt**GGTCTC**gcgttataacggcggatacg
26	*GG_Pdi*I_*Ctcel48S* _F	ttt**GGTCTC**GCCGGCcctacaaaggcacctacaaa
27	*GG_Stop_Ctcel48S* _R	ttt**GGTCTC**gcgttagttcttgtacggcaatgtatc
28	*GG_Pdi*I_*Ctcel48Y* _F	ttt**CGTCTC**GCCGGCagacaatcatccaattcaaatg
29	*GG_Stop_Ctcel48Y* _R	ttt**CGTCTC**gcgttacggttcgtttcccgatac
30	*GG_Pdi*I_*Cscel48/9A*_1F	tt**GGTCTC**GCCGGCtcgtttaactatggggaagc
31	*GG_Cs*cel48/9A_1R	tt**GGTCTC**ggcctttccatttgcataacctcagc
32	*GG_Cscel48/9A*_2F	tt**GGTCTC**aaggccttcatttaaggtcaccc
33	*GG_Cscel48/9A*_2R	tt**GGTCTC**ccttgatttggcacaaatccaccaaac
34	*GG_Cscel48/9A*_3F	tt**GGTCTC**tcaaggcctcatcacagaactgctc
35	*GG_Stop_Cscel48/9A*_3R	tt**GGTCTC**gcgttattgattaccgaacagaatttc
36	*GG_Pdi*I_*Tmcel12A*_F	ttt**GGTCTC**GCCGGCgtggtactgatgacaaaacc

37	*GG_Stop_Tmcel12A*_R	ttt**GGTCTC**gcgttagtggtactgatgacaaaacc
38	*GG_Pdi*l_*Tmcel12B*_F	ttt**GGTCTC**GCCGGCgttggtgcaacggacatttc
39	*GG_Stop_Tmcel12B*_R	ttt**GGTCTC**gcgttagttggtgcaacggacatttcc
40	*GG_Pdi*l_*Tmcel5A*_F	ttt**GGTCTC**GCCGGCggtgttgatccttttgaaagg
41	*GG_Stop_Tmcel5A*_R	ttt**GGTCTC**gcgttattcaatgctatctcctccta
42	*GG_Pdi*l_*Ctcel9I*_F	ttt**GGTCTC**GCCGGCctatggggtaggaagacctc
43	*GG_Stop_Ctcel9I*_R	ttt**GGTCTC**gcgttattatggctcttttccgtac
44	*GG_Pdi*l_*Cscel5SLH*_F	ttt**GGTCTC**GCCGGCaatactgcgtatgaaaaggataag
45	*GG_Stop_Cscel5SLH*_R	ttt**GGTCTC**gcgttacatctttcctgtaagttctaaaattttg
46	*GG_Pdi*l_*LPMO10A*_F	ttt**GGTCTC**GCCGGCggaagcgttatcaatccggc
47	*GG_Stop_LPMO10A*_R	ttt**GGTCTC**gcgttacacaaaattgacatcgctgc
48	*GG_Pdi*l_*LPMO10B*_F	ttt**GGTCTC**GCCGGCggagcgatgacgtatccgcc
49	*GG_Stop_LPMO10B*_R	ttt**GGTCTC**gcgttacgccacgctgcataagcttc
50	*GG_Pdi*l_*CcYoaJ*_F	ttt**GGTCTC**GCCGGCacaccaaaacctattccgtc
51	*GG_Stop_CcYoaJ*_R	ttt**GGTCTC**gcgttattattcagggaactggacat
52	*GG_Pdi*l_*CtPinB*_F	ttt**GGTCTC**GCCGGCtgttcaagggaaaaaaagg
53	*GG_Stop_CtPinB*_R	ttt**GGTCTC**gcgttattaatatttttcacaatcatacaa
54	*GG_Pdi*l_*Ctcel8A*_F	ttt**GGTCTC**GCCGGCggtgtgcctttaacacaaaatacc
55	*GG_Stop_Ctcel8A*_R	ttt**GGTCTC**gcgttaataaggtaggtggggtatgctc

56	*GG_Pdi*l_*Ctcel9D*_F	ttt**GGTCTC**cccggggaaaccaaagtgtcagctgc
57	*GG_Stop_Ctcel9D*_R	ttt**GGTCTC**gcgttatattggtaatttctcgattacc c
58	*GG_Pdi*l_*Cscel5*_F	ttt**GGTCTC**GCCGGCaacaaattaccgcgctac
59	*GG_Stop_Cscel5*_R	ttt**GGTCTC**gcgttattaaaagcgcataagtaac
60	*Bss*HII_AmpR_F	gcggcgcgccagttaccaatgcttaatca
61	*Zra*I_AmpR_R	gagacgtcaggtggcacttttcggggaaa
62	*GG_Ngo*MIV_SigPep1_R	ttt**GGTCTC**GCCGGCtgcagcagatgtgtcattcac
63	*GG_Ngo*MIV_Stop_F	t**GGTCTC**taacgccccgaaaggggcgtttttttgcGA GCTCgaattcgtaatcatgg
64	ClaI_SigPep02_peg2497_F	tcccctATCGATgaaacgcatgttgacagg
65	ClaI_SigPep03_peg2542_F	ttctctATCGATgcgatggatattggcagc
66	ClaI_SigPep04_C56_0963_F	ttggctATCGATgcgaatgttagccgcttt
67	ClaI_SigPep05_C56_1385_F	tttggtATCGATgaagcgaatatggctgct
68	ClaI_SigPep06_peg4352_F	tgctatATCGATgaaaacgaggtggctttt
69	ClaI_SigPep07_peg1156_F	ttgtctATCGATgaaaggatggtctaagtttttc
70	ClaI_SigPep08_peg1500_F	ttgcgtATCGATgatgaaaaaacagttatatgtatgg
71	ClaI_SigPep09_peg2593_F	aggttaATCGATgaaagtaaaggcagtagcg
72	ClaI_SigPep18_peg3497_F	tctcttATCGATgaaaaaatggaaaaaaacagc
73	ClaI_SigPep26_peg3580_F	ggagagATCGATgaaaaaatggaaatggtacc
74	ClaI_SigPep27_peg3363_F	tgccatATCGATgaaaagaaggaaagtaatggc
75	ClaI_SigPep28_C56_2367_F	accaaaATCGATgataaaaaaaacgtttgcg

76	ClaI_SigPep29_peg4265_F	ttagctATCGATgaatgcggtaaaagccac
77	ClaI_SigPep30_peg4045_F	cgacaaATCGATgtgtttgcggcggatgag
78	ClaI_SigPep31_peg1168_F	tcgcatATCGATgttgtcattttataaaaaaataactg
79	ClaI_SigPep32_peg3495_F	tcatttATCGATgcgcaggcaattggtcctg
80	ClaI_SigPep33_C56_2494_F	aaatgaATCGATgctacagtcacgaatcatc
81	ClaI_SigPep38_C56_0418_F	ctcgagATCGATgaagggtagaagacgtcttatg
82	ClaI_SigPep40_C56_0921_F	tcggttATCGATggcggagaagagaaaatttttatg
82	ClaI_SigPep41_peg0681_F	aaaccaATCGATgaacaaaacaaaaagttatttatcg
83	ClaI_SigPep42_peg3431_F	tggggtATCGATgcggataggagtacaaataag
84	ClaI_SigPep43_peg1113_F	aagagaATCGATgaaaaaatggattttggc
85	ClaI_SigPep44_C56_2000_F	caccgcATCGATgaaactgccaaaatggttg
86	ClaI_SigPep45_C56_0292_F	tctagaATCGATgcgccttttttaaatttgc
87	PdiI_SigPep02_peg2497_R	aagagGCCGGCagccatagcgggaaag
88	PdiI_SigPep03_peg2542_R	aatagGCCGGCagcagcggaagcagaaatgg
89	PdiI_SigPep04_C56_0963_R	aacagGCCGGCcgcaaatgcagcttgga
90	PdiI_SigPep05_C56_1385_R	ccaagGCCGGCagcgttagcatgcggtgg
91	PdiI_SigPep06_peg4352_R	aaaagGCCGGCttgggccgccaatgttcct
92	PdiI_SigPep07_peg1156_R	ccgagGCCGGCctgcgcctgaacaggtaaatg
93	PdiI_SigPep08_peg1500_R	cctagGCCGGCttcagcactggcgtgccatggc
94	PdiI_SigPep09_peg2593_R	cccagGCCGGCagaagcttgtaccacaaatgg
95	PdiI_SigPep18_peg3497_R	cacagGCCGGCggacgcttgggcaaagctag

96	PdiI_SigPep26_peg3580_R	cttagGCCGGCttttgcatttgctggtaaaag
97	PdiI_SigPep27_peg3363_R	attagGCCGGCtaccgcgtccgcaaaatgc
98	PdiI_SigPep28_C56_2367_R	gttagGCCGGCcgcatctgtcgaataag
99	PdiI_SigPep29_peg4265_R	tggagGCCGGCcgctgcataagtaccgg
100	PdiI_SigPep30_peg4045_R	*tttagGCCGGCcgcaagcaccggcagcac*
101	PdiI_SigPep31_peg1168_R	cggagGCCGGCatgtgcttgcggcgaagtcc
102	PdiI_SigPep32_peg3495_R	aggagGCCGGCgctcgcttctgcttgctcgg
103	PdiI_SigPep33_C56_2494_R	gggagGCCGGCctccgcaaaaacaggaaactc
104	PdiI_SigPep38_C56_0418_R	agtagGCCGGCttccgcgtaaaccgtcattc
105	PdiI_SigPep40_C56_0921_R	tgaagGCCGGCcccaaacgctacagca
106	PdiI_SigPep41_peg0681_R	ctgagGCCGGCttgcgcttgggcaattcctg
107	PdiI_SigPep42_peg3431_R	actagGCCGGCggcatatgccggagaaatg
108	PdiI_SigPep43_peg1113_R	cagagGCCGGCatcggcaaaaacattctccg
109	PdiI_SigPep44_C56_2000_R	gtgagGCCGGCggccattaatgaagcc
110	PdiI_SigPep45_C56_0292_R	tcaagGCCGGCtgccatcacggatagag

Table 20 - Plasmids constructed or obtained for this study

Plasmid	Size (bp)	Reference
pUCG18	7,490	(Taylor *et al.*, 2008)
pUCG3.8	5,244	(Bartosiak-Jentys *et al.*, 2013)
pUCG4.8	4,758	This study
pUCG18_pRPLS_sfGFP	-	Ben Reeve, Imperial College, UK
pUCG18_pUP2n38_sfGFP	-	Ben Reeve, Imperial College, UK
pUCG4.8_GoldenGate_sfGFP	5,496	This study
pUCG4.8_pRPLS_sfGFP	5,710	This study
pUCG4.8_pUP2n38_sfGFP	5,566	This study
pUCG4.8_p4070_sfGFP	5,937	This study
pUCG4.8_pAdhE_sfGFP	5,909	This study
pUCG4.8_pTre_sfGFP	5,739	This study
pUCG4.8_pSorb_sfGFP	5,783	This study
pUCG3.8_pRPLS_sfGFP	4,726	This study
pUCG4.8_pRPLS_SP1_GoldenGate	4,960	This study
pUCG4.8_pUP2n38_SP1_GoldenGate	4,816	This study
pUCG4.8_pRPLS_SP1_*Tf*cel6B	6,r817	This study
pUCG4.8_pRPLS_SP1_*Ct*cel48S	7,228	This study
pUCG4.8_pRPLS_SP1_*Ct*cel48Y	7,783	This study
pUCG4.8_pRPLS_SP1_*Tm*cel12A	5,860	This study

pUCG4.8_pRPLS_SP1_*Tm*cel12B	5,836	This study
pUCG4.8_pRPLS_SP1_*Tm*cel5A	6,037	This study
pUCG4.8_pRPLS_SP1_*Ct*cel9I	7,660	This study
pUCG4.8_pUP2n38_SP1_*Tm*cel12A	5,716	This study
pUCG4.8_pUP2n38_SP1_*Tm*cel12B	5,692	This study
pUCG4.8_pRPLS_SP1_*Ct*cel8A	6,421	This study
pUCG4.8_pRPLS_SP1_*Ct*cel9D	6,910	This study
pUCG4.8_pRPLS_SP1_*Cs*cel5	6,070	This study
pUCG4.8_pUP2n38_SP1_*CcYoaJ*	5,604	This study
pUCG4.8_pUP2n38_SP1_*Ct*PinA	6,103	This study
pUCG4.8_pRPLS_SPL1.0	-	This study
pUCG4.8_pRPLS_SPL1.0_*Tm*cel12B	-	This study

APPENDIX II

Table 21 – Published biochemical properties and specific activities of characterised bacterial endo- and exoglucanases. Optimal temperatures (T_{opt} in °C), optimal pH, and specific activities (U/mg) of exoglucanases reported on multiple cellulosic substrates.

Organism	Cellulase Name	pH	T_{opt}	CMC (U/mg)	Avicel (U/mg)	Bglucan (U/mg)	PASC (U/mg)	Description of published cellulolytic activity	Reference
***Geobacillus* sp. 70PC53**	CelA	5	65	116.4	0	1267.3	41.4	(pH 5 & 65°C) 0 U/mg (Avicel), 41.4 U/mg (Acid swollen Avicel), 116.4 U/mg (CMC), 0 U/mg (Cellulose fiber), 1,267.3 U/mg (b-D-Glucan (barley)), 1.0 U/mg (Filter paper), 945.4 U/mg (Lichenan), 5.3 U/mg (Xylan-birchwood), 0.1 U/mg (Xylan oat spelts)	I-Son Ng et al, 2009
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [wt]	6.5	75	294	0	0	0	(pH 6.5 & 75°C) 294 U/mg (0.4%w/v CMC)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [3F6]	6.5	75	396	1.5	1743	0	(pH 6.5 & 75°C) 396 U/mg (0.4%w/v CMC), 1743 U/mg (β-glucan barley), 1.50 U/mg (Avicel)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [C3-13]	6.5	75	567	1.43	2689	0	(pH 6.5 & 75°C) 567 U/mg (0.4%w/v CMC), 2689 U/mg (β-glucan barley), 1.43 U/mg (Avicel)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [V249A]	6.5	75	388	0	0	0	(pH 6.5 & 75°C) 388 U/mg (0.4%w/v CMC)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [I13V]	6.5	75	289	0	0	0	(pH 6.5 & 75°C) 289 U/mg (0.4%w/v CMC)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB4**	Cel5A [I321V]	6.5	75	393	0	0	0	(pH 6.5 & 75°C) 393 U/mg (0.4%w/v CMC)	Liang et al, 2011
Alicyclobacillus acidocaldrious	Cel9A (59 kDa)	5	70	73.30508475	0.008474576	200	11		Telke et al, 2012

Alicyclobacillus acidocaldrious	CBM3–Cel9A (76 kDa)	5	70	86.842 10526	0.0789 47368	168.421 0526	13.157 89474		Telke et al, 2012
Alicyclobacillus acidocaldrious	CBM4–Cel9A (80 kDa)	5	70	91.5	0.05	187.5	15.75		Telke et al, 2012
Alicyclobacillus acidocaldrious	CBM30–Cel9A (80 kDa)	5	70	96.5	0.0625	187.5	16		Telke et al, 2012
T. reesei	Cel12A (Endo) [wt]	5	50	22	0	0	0		Nakazawa H. et al, 2009
T. reesei	Cel12A (Endo) [1R21]	5	50	19	0	0	0		Nakazawa H. et al, 2009
T. reesei	Cel12A (Endo) [1R41]	5	50	18	0	0	0		Nakazawa H. et al, 2009
T. reesei	Cel12A (Endo) [PFS]	5	50	24	0	0	0		Nakazawa H. et al, 2009
T. reesei	Cel12A (Endo) [2R4]	5	50	26	0	0	0		Nakazawa H. et al, 2009
Opuntia vulgaris	endo-1, 4-β-d-glucanase T50 isoform	7	50	125	0	0	0	(pH 7.0 & 50°C) 125 U/mg (1%w/v CMC)	Shyamala S et al, 2011
Opuntia vulgaris	endo-1, 4-β-d-glucanase T70 isoform	7	50	144	0	0	0	(pH 7.0 & 50°C) 144 U/mg (1%w/v CMC)	Shyamala S et al, 2011
Opuntia vulgaris	endo-1, 4-β-d-glucanase T90 isoform	7	50	137	0	0	0	(pH 7.0 & 50°C) 137 U/mg (1%w/v CMC)	Shyamala S et al, 2011

Clostridium thermocellum	CelA/Cthe_0269 [GH8]	6	60	580	5	2910	208	(pH 6.0 & 60°C) 580 U/mg (CMC), 5 U/mg (Avicel), 2910 U/mg (Barley β-glucan), 208 U/mg (Acid swollen Avicel)	Schwarz et al, 1986
Clostridium thermocellum	CelB/Cthe_0536 [GH5]	6.3	60	7.5	0	0	0	(pH 6.3 & 60°C) 7.5 U/mg (CMC)	Beguin et al, 1983
Clostridium thermocellum	CelC/Cthe_2807 [GH5]	6.3	60	0.16	0	0	0	(pH 6.3 & 60°C) 0.16 U/mg (CMC), 1.8 U/mg (pNPC)	Petre et al, 1986
Clostridium thermocellum	CelD/Cthe_0825 [GH9]	5	50	567.2	0	0	247.4	(pH 5 & 50°C) 567.2 U/µmol (CMC), 247.4 U/µmol (PASC), 0 U/µmol (salicin)	Lee H et al, 2010
Clostridium thermocellum	CelG/Cthe_2872 [GH5]	6.3	60	37.1	0	0	0	(pH 6.3 & 60°C) av. 37.1 U/mg (1.5%w/v CMC), Km=1.925 | Vmax= 5.275 (cPNC), Km=0.975 | Vmax=0.1275 (pNPC)	Lemaire M et al, 1993
Clostridium thermocellum	CelI/Cthe_0040 [GH9,CBM3,CBM3]	6	60	2.3	0.001	300.7	2.4	(pH 6.0 & 60°C) 2.3 U/mg (1%w/v CMC), <0.01 U/mg (2%w/v Avicel), 300.7 U/mg (2%w/v Barley β-glucan), 2.4 U/mg (2%w/v PASC)	Zverlov et al, 2003
Clostridium thermocellum	CelI/Cthe_0040 [GH9,CBM3,CBM3]	6.5	60	2.3	0.001	300.7	2.4	(pH 6.5 & 60°C) 0.8 U/mg (CMC), <0.01 U/mg (Avicel), 108.7 U/mg (Barley β-glucan), 2.4 U/mg (Lichenan)	Hazelwood G. et al, 1992
Clostridium thermocellum	CelJ/Cthe_0624 [CBM30,GH9,GH44,CBM44]	6.3	60	73.59550562	0.05505618	48.93258427	0	(pH 6.3 & 60°C) 13,100 U/µmol (1%w/v CMC), 9.8 U/µmol (1%w/v Avicel), 8,710 U/µmol (2%w/v Barley β-Glucan)	Arai T et al, 2003
Clostridium thermocellum	CelN/Cthe_0043 [GH9,CBM3]	5.4	70	3.4	0.001	360	2.6	(pH 5.4 & 70°C) 3.4 U/mg (1%w/v CMC), <0.01 U/mg (2%w/v Avicel), 360 U/mg (2%w/v Barley β-glucan), 2.6 U/mg (2%w/v PASC)	Zverlov et al, 2003
Clostridium thermocellum	CelO/Cthe_2147 [CBM3,GH5]	6.6	65	7.5	0.002	340	0	(pH 6.6 & 65°C) 7.5 U/mg (CMC), 0.002 U/mg (Avicel), 340 U/mg (Barley β-glucan)	Zverlov et al, 2002
Clostridium thermocellum	CelQ/Cthe_0625 [GH9,CBM3]	7	60	392	0	159	0	(pH 7.0 & 60°C) 392 U/mg (1%w/v CMC), 159 U/mg (Barley β-Glucan)	Kurokawa et al, 2002

Clostridium thermocellum	CelR/Cthe_0578 [GH9,CBM3]	6	78	16.66666667	0	519.8666667	2.573333333	(pH 6.0 & 78°C) 1250 U/µmol (1%w/v CMC), 193 U/µmol (1%w/v PASC), 38,990 U/µmol (2%w/v Barley β-Glucan)	Zverlov et al, 2005
Clostridium thermocellum	CelT/Cthe_2812 [GH9]	7	60	126	0	160	0	(pH 7.0 & 60°C) 126 U/mg (1%w/v CMC), 160 U/mg (Barley β-Glucan)	Kurokawa et al, 2002
***Acidothermus cellulolyticus* 11B ATCC 43068**	Acel_0614 EGAc;EGI (GH5,CBM2) (endo-β-1,4-glucanase E1)	5	50	40	0	0	0	(pH 5 & 50°C) 40 U/mg (5 g/L CMC)	0
***Anaerocellum thermophilum* DSM 6725**	Athe_1867 (CelA)(GH9,CBM3, CBM3,CBM3,GH48) (Exoglucanase)	6	72	1.758	0.055	100.51	0.082	(pH 6.0 & 72°C) 1.758 U/µmol (0.5%w/v CMC), 0.055 U/µmol (1%w/v Avicel), 0.082 U/µmol (1%w/v Acid-swollen Avicel), 100.510 U/µmol (1%w/v Barley β-Glucan)	Zverlov et al, 1998
***Anaerocellum thermophilum* DSM 6725**	Athe_1867 (CelA-Truncated) (Exoglucanase)	6	72	1.801	0.018	88.244	0.076	(pH 6.0 & 72°C) 1.801 U/µmol (0.5%w/v CMC), 0.018 U/µmol (1%w/v Avicel), 0.076 U/µmol (1%w/v Acid-swollen Avicel), 88.244 U/µmol (1%w/v Barley β-Glucan)	Zverlov et al, 1998
***Caldicellulosiruptor saccharolyticus* DSM 8903**	Csac_0678 (GH5,CBM28) (Endoglucanase with SLH DOMAINS)	6.5	60	8.94	0.0113	28.2	0	(pH 6.5 & 60°C) 8.94 U/mg (CMC), 5.52 U/mg (p-Nitrophenyl cellobioside), 0.0113 U/mg (Avicel), 28.2 U/mg (Barley B-Glucan)	Ozdemir I et al, 2012
***Thermobifida fusca* YX**	TFU_1627 (E1;EGtf1) (CBM4,GH9,CBM2) (endo-?-1,4-glucanase)	7	65	768	0	0	87	(pH 7.0 & 65°C) 768 U/mg (1%w/v CMC), 87 U/mg (1%w/v PASC), 15.8 U/mg (2%w/v Filter Paper)	Calza R et al, 1985
***Thermotoga maritima* MSB8**	TM1524 (CelA;TmCel12A)	6	85	302	0	1785	0	(pH 6.0 & 85°C) 302 U/mg (CMC), 0.00 U/mg (Avicel), 1785 U/mg (Barley β-glucan)	Liebl W et al, 1996

	(GH12)(endo-beta-1,4-glucanase A)								
***Thermotoga maritima* MSB8**	TM1525 (CelB) (GH12)(endo-beta-1,4-glucanase B)	6	85	0.89	0	2.905	0	(pH 6.0 & 85°C) 0.890 U/mg (CMC), 0.00 U/mg (Avicel), 2.905 U/mg (Barley β-glucan)	Liebl W et al, 1996
***Thermotoga maritima* MSB8**	TM1751 (GH5)(endoglucanase - or maybe mannanase?)	7	80	616	0	2345	0	(pH 7.0 & 80°C) 616 U/mg (CMC), 2345 U/mg (Barley β-glucan)	Chhabra R et al, 2002
***Thermoanaerobacter tengcongensis* MB5**	Cel5A [3F6]	6.5	75	396	1.5	1743	0	(pH 6.5 & 75°C) 396 U/mg (0.4%w/v CMC), 1743 U/mg (β-glucan barley), 1.50 U/mg (Avicel)	Liang et al, 2011
***Thermoanaerobacter tengcongensis* MB6**	Cel5A [C3-13]	6.5	75	567	1.43	2689	0	(pH 6.5 & 75°C) 567 U/mg (0.4%w/v CMC), 2689 U/mg (β-glucan barley), 1.43 U/mg (Avicel)	Liang et al, 2011
Caldibacillus cellulovorans	CMCase (Exo)	7	70	30.6	1.7	0	0	(pH 7.0 & 70°C) 30.6 U/mg (CMC), 1.7 U/mg (Avicel)	Huang et al, 2003
***Clostridium phytofermentans* ISDg**	Cel48 (Exo)	6	50	0.0475	0.1875	0	0	(pH 6.0 & 50°C) 3.8±0.2 U/µmol (CMC), 15.1±0.6 U/µmol (Avicel)	Zhang et al, 2010
Clostridium stercorcarium	Avicelase I	6	70	11.5	0.302	481.23	4.09		Bronnenmeier K et al, 1991
Clostridium stercorcarium	Avicelase II	6	70	0	0.041	0	0.102		Bronnenmeier K et al, 1991
Clostridium thermocellum	CelF/Cthe_0543 [GH9,CBM3]	6	45	0	0.163414634	0	0.518292683	(pH 6.0 & 45°C) 0 U/µmol (CMC), 42.5 U/µmol (PASC), 13.4 U/µmol (Avicel), 3.9 U/µmol (BMC-cellulose)	Reverbel-Leroy C et al, 1997
Clostridium thermocellum	CelK/Cthe_0412 [CBM4,GH9]	6	60	0	0	0	0	(pH 6.0 & 60°C) Km=1.67 µM | Vmax=15.1 U/mg (pNPC)	Kataeva I et al, 1999

Clostridium thermocellum	CelS/Cthe_2089 [GH48]	5.7	60	8	19	0	0	(pH 5.7 & 60°C) 8 U/mg (CMC), 19.0 U/mg (Avicel), 150 U/mg (Amorphous cellulose)	Kruus et al, 1995
Clostridium thermocellum	CelY/Cthe_0071 [GH48,CBM3]	7	65	0.17	0.25	15.8	29.9	(pH 7.0 & 65°C) 0.17 U/mg (1%w/v CMC), 0.25 U/mg (2%w/v Avicel), 15.8 U/mg (2%w/v Barley β-glucan), 29.9 U/mg (2%w/v PASC)	Berger et al, 2007
Clostridium thermocellum	Cthe_0413(CbhA)[CBM4,GH9,CBM3](cellobiohydrolase A -non-reducing end)	5	50	231.8	0	0	102.7	(pH 5 & 50°C) 231.8 U/µmol (CMC), 102.7 U/µmol (PASC), 0 U/µmol (salicin)	Lee H et al, 2010
Clostridium thermocellum	Cthe_1256[GH3](exo-/cellobiohydrolase)	6.5	60	18.8	2.1	0	0	(pH 6.5 & 60°C) 18.8 U/mg (CMC), 2.1 U/mg (Avicel), 18.7 U/mg (pNPC), 24.6 U/mg (Lichenan)	Singh R et al, 1993
Clostridium thermocellum	S8-tr	5	60	0.28	0.18	0	0	(pH 5.0 & 60 °C) 0.028 U/mg (CMC), 2.6 U/mg (Amorphous cellulose), 0.18 U/mg (Microcrystalline cellulose)	Morag et al, 1990
***Thermobifida fusca* YX**	TFU_1074 (E2;CelB) (GH6,CBM2) (1,4-beta-cellobiosidase A)	7	65	0.17	0.25	15.8	29.9	(pH 7.0 & 65°C) 0.17 U/mg (1%w/v CMC), 0.25 U/mg (2%w/v Avicel), 15.8 U/mg (2%w/v Barley β-glucan), 29.9 U/mg (2%w/v PASC)	Calza R et al, 1985